SPEECH MONOGRAPHS

Published by

THE SPEECH ASSOCIATION OF AMERICA

Volume XVIII—No. 4 November, 1951

SPEECH MONOGRAPHS

VOLUME XVIII, No. 4 November, 1951

AN EXPERIMENTAL STUDY OF CERTAIN PHYSIOLOGICAL, INTROSPECTIVE AND RATING-SCALE TECHNIQUES FOR THE MEASUREMENT OF STAGE FRIGHT

MILTON DICKENS, *University of Southern California*
WILLIAM R. PARKER, *Redlands University*

I. The Problem

Experimental research on stage fright phenomena will be seriously limited unless and until techniques, of satisfactory validity and reliability, can be devised for the measurement of those phenomena. For example, Holtzman[1] in a recent study undertook to explore possible differences in "personality structure" among persons manifesting various degrees of stage fright. To measure personality factors, Holtzman used the Minnesota Multiphasic Personality Inventory (MMPI); while to measure stage fright, he used three techniques: two subjective inventories and a judges' rating scale. Commenting upon these stage fright measuring techniques, Holtzman said:

> In a real sense, these are rulers marked off in identical terms but with very different calibrations. This became clear when time-consuming attempts were made, early in the study, to deal only with subjects on whom the various measures of stage fright tended to agree. An obviously false and unworkable distribution resulted.[2]

Using 498 men and women students at San Francisco State College for subjects, Holtzman found that personality patterns, as measured by MMPI, varied according to which of his three measures of stage fright was used; and that his data were further confused by sex differences. He concluded:

> Any final statement in terms of a general concept of stage fright is not warranted by the scope or data of this study. On the contrary, it can only be noted again that stage fright defies any but operational definition and that its relationships to personality structure depend upon the nature of that definition.[3]

The basic importance of developing measuring techniques may be further clarified when it is considered that experimentation with stage fright *therapies* cannot proceed beyond a rule-of-thumb phase unless and until it becomes possible to measure stage fright before and after the controlled application of various such therapies.

Theoretically, there appear to be three possible types of measuring techniques or indices, appropriate to the phenomena of stage fright: (1) introspective reports, (2) reports by observers, and (3) physiological changes.

Several studies have been reported involving the first of these approaches. The usual procedure has been to administer pencil-and-paper questionnaires, aimed at securing systematic introspec-

[1] Paul D. Holtzman, *An Experimental Study of Some Relationships Among Several Indices of Stage Fright and Personality*, Ph.D. Dissertation, University of Southern California, 1950.

[2] *Ibid.*, 4.

[3] *Ibid.*, 55.

tive reports in a form which can be expressed in quantitative terms. The most thorough study of this type appears to be that of Gilkinson[4] in which a Personal Report of Confidence of Speakers (PRCS), comprising 104 items expressing feelings of confidence or fear, was given to 420 men and women speech students at the University of Minnesota. Gilkinson reported a satisfactory degree of statistical reliability on the basis of internal consistency of the PRCS items. He did not, however, validate the PRCS against any direct and independent criterion.

Dickens, Gibson and Prall[5] recently described experiments in which the second possible type of measurement, reports by observers, was emphasized. In these experiments, 40 male student speakers at the University of Southern California were rated for "observable degrees of stage fright" by 61 "expert" judges (speech teachers and graduate students) on a five-step rating scale. Sound motion pictures and Gilkinson's PRCS were additional techniques. Dickens, Gibson and Prall found a correlation of $+.59 \pm .104$ between their PRCS scores and their judges' ratings (JR's). This correlation seemed reasonable in view of the fact that PRCS purported to measure how the speakers *felt*, while JR purported to measure how they *looked and sounded*. It was concluded that PRCS and JR tended to validate one another. The experimenters also reported that a split-half comparison of the 61 expert judges showed the JR technique to be remarkably stable; that the technique remained highly reliable when ratings of as few as five judges were averaged; that judges' ratings appeared to be more accurate when judging sound motion pictures than when judging the speakers "in the flesh."

The present study sought to emphasize the third possible type of stage fright measurement, physiological changes. The principal problems were:

(1) How much physiological disturbance occurs in students in a classroom public speaking situation, as indicated by pulse and blood pressure readings taken immediately before and immediately after speaking?

(2) What inter-relationships are found between these pulse and blood pressure fluctuations and (a) introspective reports, as indicated by Gilkinson's PRCS inventory, and (b) observers' reports, as indicated by a rating scale marked by speakers' classmates?

(3) What sex differences, if any, are found when the foregoing techniques for stage fright measurement are applied?

II. Subjects, Materials, Procedures

In general, experimental procedures were as follows. One hundred college students, 50 male and 50 female, gave regularly assigned classroom speeches. Each was rated by his classmates for "observable degrees of stage fright." Immediately following his speech, each went into an anteroom where pulse and blood pressure readings were taken. Each speaker then filled out a copy of Gilkinson's PRCS inventory. Approximately a month later, the same procedure was repeated with the important difference that pulse and blood pressure readings were secured immediately before, rather than after, the speech. These procedures will next be described in greater detail.

Speakers. All subjects were members of the Fundamentals of Speech course

[4] Howard Gilkinson, "Social Fears As Reported by Students in College Speech Classes," *Speech Monographs*, IX (1942), 141-160.

[5] Milton Dickens, Francis Gibson and Caleb Prall, "An Experimental Study of the Overt Manifestations of Stage Fright," *Speech Monographs*, Vol. 17, No. 1 (1950).

which is required for graduation at Redlands University. There were approximately 150 students in the course at the time of this study. To facilitate statistical calculations and comparisons, 50 males and 50 females were chosen at random as the subjects. Most of these were Freshmen or Sophomores. Average age for the women was nineteen; for the men, twenty-two.

Judges. Each speaker was rated by his classmates. The class sections ranged in size from twelve to sixteen. There were never fewer than ten judges in a class at any one time; the average per meeting was thirteen.

Rating Scale. Judges were provided with mimeographed booklets of the rating scale developed by Dickens, Gibson and Prall. This is a linear scale of five steps: (1) Virtually no observable degree of stage fright, (2) Less than average degree of stage fright, (3) Average degree of stage fright, (4) More than average degree of stage fright, (5) Extreme degree of stage fright.

Inventory. Speakers were provided with mimeographed copies of Gilkinson's PRCS inventory, comprising 104 statements which may describe a speaker's feeling. The first 54 of these statements are considered "negative" or "fear" items; the remaining 50 are considered "positive" or "confidence" items. Subjects may mark each item as "yes," "?" or "no."

Physiological Measurements. A registered nurse, qualified to measure blood pressure, and another nurse to take pulse readings, were employed. A graduate assistant helped them with their instruments. The instruments included a watch with a large sweep second hand, a sphygmomanometer, a stethoscope, and lists prepared for convenient recording. For each class meeting the list contained the names of students who were to speak that hour; a single column was provided in which to record pulse; a double column in which to record systolic and diastolic blood pressure.

Procedure. Approximately two months after the beginning of the semester, all Fundamentals of Speech sections were given the regular assignment of a three to five minute, extemporaneous talk. During the experimental sessions, there was minimum disturbance of the customary class routine. Students were provided with copies of the rating scale, together with instruction sheets, and were asked to rate each speaker. These judges' ratings (JR's) were collected by the instructor at the end of each speech and placed in an envelope bearing the particular speaker's name. Each speaker was called upon by his instructor to give his talk and at the conclusion, instead of going back to his seat, was asked to step into the adjoining anteroom for further assignment before rejoining the class.

The door to the anteroom was located just a few steps from the speaking area. Use of a stop watch established that in no case did more than 30 seconds elapse between the ending of the speech and the beginning of the testing in the anteroom.

When a speaker stepped into the anteroom, he was informed that pulse and body pressure readings were to be taken. The term, "body pressure," was used throughout the experiments since, according to consulting physicians, the word, "blood," is emotionally loaded for some persons and may itself cause pulse or body pressure fluctuations.

One nurse immediately applied the apparatus for measuring body pressure, and simultaneously the other nurse took the pulse reading. Data were at once recorded on the prepared sheets.

Subjects remained standing for these tests. Since they were standing during

their speeches, it appeared wise to take pulse and body pressure readings in the same position. Medical authorities point out that posture may affect blood pressure. For example, Best and Taylor[6] state that,

> The diastolic pressure is somewhat higher in the standing position than in the sitting position and lowest in recumbency. This change is found to occur whether the postural change is brought about actively or passively and is evidently an over-compensation for the gravity effect. The systolic pressure usually rises but to a less extent than the diastolic, so the pulse pressure is reduced.

Following the physiological readings, the student was handed a copy of the PRCS inventory which he filled out while in the anteroom. Upon completion of the PRCS, he returned to his class to participate in judging the other speakers, using the JR scales.

All the foregoing procedures were applied until 100 subjects had been tested. Taken together they were called Session One. About one month later, Session Two was accomplished. Procedures in Session Two varied from Session One in that pulse and body pressure data were secured immediately *before* the student gave his speech. Thus, when a student was called on, he arose from his chair and went directly to the anteroom where the readings were taken. He proceeded immediately to his place in front of the class and commenced his speech. At the conclusion of his talk, he returned to his seat and filled out a copy of the PRCS, after which he resumed his duties as judge of the other student speakers.

In order to study the possible fluctuations accompanying the experimental situations, it was necessary, of course, to establish "normal" pulse and body pressure reading for each subject. Pulse and body pressure were known to be affected by a variety of factors: recency of eating, minor illnesses, bad news from home, etc. Therefore, after consultation with physicians, it was decided to establish normal pulse and body pressure by averaging several readings taken under reasonably ordinary conditions. Students were asked to report to the testing room at the same hour as their speech class but on several different days. From three to five such readings were eventually secured for each subject. Resulting averages were treated as the "normal" readings.

In setting up the several details of the foregoing experimental design, it was necessary, of course, to make a number of choices. Thus, in attempting to measure introspective reports of speakers, the question was whether to use PRCS or to devise, and undertake to validate, a new inventory. In favor of the former course was the fact that PRCS had been used with reasonable success in at least three previous studies; in favor of the latter course, however, was the fact that previous studies had revealed several weaknesses in PRCS. In attempting to secure observers' ratings, the question was whether to try to secure expert judges or to use students' classmates. Here practical considerations weighed heavily: it would have been exceedingly difficult to muster a suitable number of speech teachers and graduate students. Still more decisive was the belief that ten to fifteen student judgments averaged would prove about as valid and reliable as five or six experts.

Other choices had to be made with regard to physiological techniques. Since the early classic research of W. B. Cannon and his students, much has been learned of the diverse physiological changes which often occur under conditions of emotional stress: adrenal glands

[6] Charles H. Best and Norman B. Taylor, *The Physiological Basis of Medical Practice* (Baltimore: The Williams and Wilking Company, 1945), 128.

secrete into the blood; digestion halts; liver releases stored sugar; pulse and blood pressure fluctuate; breathing is affected; pupil of eye enlarges; sweat breaks out; skin temperature rises or falls; etc. It was apparent that all these various possibilities could not be measured in connection with an ordinary classroom speech. The question was: Which ones? Careful attention was given the possibilty of taking blood samples and securing analyses of blood content. This was eventually rejected on several grounds: (a) university authorities doubted that blood samples should be taken without written consent of students; (b) for an indeterminate number of subjects, we would probably have been measuring emotional disturbances at the prospect of blood-giving rather than speech-giving; (c) laboratory physicians were dubious as to whether measurable differences should be expected under the described experimental setup; (d) laboratory costs for blood analyses would be extremely high. In similar fashion, several other possible kinds of physiological measurement were considered and rejected.

The decision to try pulse and blood pressure data was based on several grounds. Securing the readings would be relatively simple, quick and inexpensive. Reading could be taken within 30 seconds of the beginning or end of a speech, a procedure which the experimenters considered advantageous as compared with taking physiological data during the speech since the latter procedure would involve greater danger of perverting the experimental situation. Finally, it was hoped that the three types of readings, taken together, might give a fairly accurate measure of overall circulatory disturbance which, in turn, might suffice as a gross index to general physiological disturbance during emotional states.

III. Presentation and Analysis of Data

During the course of the experiments, approximately 4,600 scores were obtained. These have been combined into 13 sets of means, standard deviations, and standard errors of the means for each sex in Table I. The table presents introspective scores (PR); averaged judges' ratings (JR); pulse; systolic blood pressure (Syst); and diastolic blood pressure (Dia); covering two experimental sessions (1 and 2) and the normal readings (N).

TABLE I

Means, Standard Deviations, and Standard Errors of Means for 50 Male and 50 Female Subjects, Measuring Aspects of Stage Fright: Introspective Inventory (PR), Averaged Judges' Ratings (JR), Pulse, Systolic and Diastolic Blood Pressure: Covering Two Experimental Sessions (1 and 2) and Normal Readings (N).

	Males			Females		
	M	σ	σ_M	M	σ	σ_M
PR-1	43.2	17.7	2.5	50.4	16.0	2.3
PR-2	40.96	16.3	2.3	44.7	13.7	2.0
JR-1	2.46	.71	.1	1.97	.46	.07
JR-2	2.77	.56	.08	2.47	.42	.06
Pulse-N	76.5	6.9	.99	75.5	5.8	.8
Pulse-1	86.1	17.4	2.5	92.4	13.3	1.9
Pulse-2	95.0	15.5	2.2	96.0	13.8	2.0
Syst-N	120.9	8.7	1.2	117.9	6.2	.9
Syst-1	129.5	12.4	1.8	119.6	14.0	2.0
Syst-2	131.9	10.5	1.5	121.0	8.7	1.2
Dia-N	76.0	5.9	.08	72.0	5.7	.8
Dia-1	77.3	10.1	1.4	73.9	7.0	1.0
Dia-2	78.9	7.4	1.1	74.5	7.1	1.0

In analyzing the foregoing data, the experimenters were first concerned with the question of whether or not there were notable changes in pulse and blood pressure under the experimental conditions. Of the 600 instances where comparisons could be made between normal and experimental readings of pulse, systolic and diastolic blood pressure, 565 showed differences and 35 remained the same. Thus, with comparative percentages of approximately 94 to 6, there was no doubt as to the *fact* of change.

The second question was regarding the *amount* of change. The means in Table I conceal the total amount of fluctuation because in some cases the pulse or blood pressure went up, while in others, down. Disregarding the direction of changes, the means for amounts of fluctuation for all 100 subjects in both sessions were: 18.1 for pulse; 10.1 for systolic; and 7.3 for diastolic. These figures indicated that the effects of even a "routine" class speech upon students' pulse rates and blood pressures were pronounced, and challenged explanation.

A third question concerned the *direction of changes.* Of the 565 instances of fluctuation, 412 were greater than the normal, and 153 were less. Additional evidence was provided when appropriate mean differences were computed and evaluated.[7] Inspection of means in Table I revealed that all the normal physiological means increased under experimental conditions in both Sessions One and Two. This tendency to rise was most marked in the case of pulse means: all four of the appropriate mean pulse comparisons revealed very significant differences. Similarly, the four comparisons for systolic means showed two very significant; one significant; and one not significant. However, the four comparisons for diastolic showed one significant and three not significant. Taken together, these data led to the conclusion that circulatory rates tended to increase significantly from normal although in a substantial minority of instances the rates decreased.

The next comparison was between physiological fluctuations in Sessions One and Two. Subjects were classified in terms of whether amounts of fluctuation were greater in Session One; greater in Session Two; or the same for both. Chi square tests for significance[8] were applied. Results were as follows: (1) a very significantly larger number of subjects showed greater pulse fluctuation before speaking; (2) an insignificantly larger number of subjects showed greater systolic fluctuation before speaking; (3) an insignificantly larger number of subjects showed greater diastolic fluctuation after speaking. These findings throw some light upon the often voiced hypothesis that in most people the anticipation of speaking is more disturbing than the act itself. Apparently, interpretation of these data should be modified by consideration of the fact that, of the three physiological measures used, pulse rate is most immediately sensitive to stimuli[9] while diastolic pressure may be expected to respond more slowly but more fundamentally.[10] Of course, these data should be considered as suggestive

[7] In comparing means, customary statistical formulas were used by which the standard errors of the differences between means, and the *t*-ratios, were computed: J. P. Guilford, *Fundamental Statistics in Psychology and Education,* 2nd ed. (McGraw-Hill, New York, 1950), 213 and 216. Significance of *t*-ratios was estimated from Wallace and Snedecor tables: Guilford, *op. cit.,* 609-610. To indicate significance in this paper, "very significant" (VS) will mean significant at the 1% level of confidence; "significant" (S) will mean significant at the 5% level; and "not significant" (NS) will mean below the 5% level.

[8] Guilford, *op. cit.,* 273 ff.

[9] See, for example: William Zoethout and W. W. Tuttle, *Textbook of Physiology,* 10th ed. (St. Louis: C. V. Mosby Co., 1949), 203 ff.

[10] *Ibid.,* 195 ff.

only, since two speeches were given in Sessions One and Two; more decisive data must await experiments in which measurements are made immediately before and after the same speech performance.

The experimenters wondered if their data would show any significant differences among the three physiological measures in terms of the variation of the fluctuations. Therefore, standard deviations for pulse, systolic and diastolic were compared, for both sexes and both experimental sessions. None of these differences were statistically significant.

Turning to the second major part of the problem, relationships between the physiological scores and the PRCS scores and the JR scores were considered. To make possible some of the combinations and comparisons, discussed below, all data were translated into *T*-scores by means of McCall's *T*-scale technique.[11] All correlation coefficients were computed by the Pearson product-moment method.[12]

The first main issue was, of course: to what extent did the physiological scores, introspective reports, and judges' ratings tend to validate one another? It soon became apparent that no simple answer to this question could be given. A large number of correlations and other statistical manipulations were tried and discarded. Frequent choices, both as to techniques and interpretations, had to be made.

Thus, for example, a series of difficulties arose in trying to decide upon the most appropriate way by which to express the physiological scores. Should the original scores for pulse and blood pressure rates be used? Or the differences between normal and experimental rates? Or a ratio between normal and experimental? Should the three physiological measures be treated separately or in combination? If combined, should the three be given equal weight? Should the systolic and diastolic measures be translated into "mean pressures?"[13] How were cases to be interpreted where pulse or blood pressure dropped below normal? If no change in score was an indication of "confidence," was a negative score "super-confidence?" Or did it indicate even greater fear than a similar amount of rise in rate? These, and other, questions were encountered, explored statistically, and tentatively resolved.

The most defensible treatment of the physiological data appeared to be as follows: (a) calculate the differences between the normal and experimental rates, regardless of sign, i.e., consider the *amount* of fluctuation from normal, rather than its direction, as the best index of emotional disturbance; (b) translate fluctuations of pulse, systolic and diastolic into *T*-scores to make them comparable and then average the three without weighting, i.e., consider the three combined as the best index of "general circulatory disturbance"; (c) average the resulting "physiological" scores for Session One (after speaking) and Session Two (before speaking), i.e., consider that some persons respond more to the anticipation of speaking and others to the act itself, and that the best over-all index would reflect either or both.

In order to match the foregoing treatment of physiological scores, it was decided to also average PR'S and JR's for Sessions One and Two.[14]

[11] Guilford, *op. cit.*, 296 ff.
[12] *Ibid.*, 157 ff.

[13] Zoethout and Tuttle, *op. cit.*, 183: "The difference between these two pressures is termed *pulse pressure*. The *mean pressure* is usually stated as the diastolic pressure plus from one-third to one-half of the pulse pressure."

[14] To indicate these several combinations of data, the following set of symbols will be used. To show which sessions, "1" will symbolize Session One; "2" will symbolize Session Two;

The correlation coefficients resulting after the foregoing treatment of data for males only, are presented as Table II. It was seen that, of the 27 *r* values, all but one were positive. Applying a null hypothesis and utilizing Wallace and Snedecor tables, it was noted that twelve of the coefficients were very significant and six were significant, for a total of eighteen, as opposed to six which were not significant. The three most crucial values were:

PR-12 *vs* JR-12	.46 (VS)
Phys-12 *vs* PR-12	.33 (S)
Phys-12 *vs* JR-12	.64 (VS)

Considered as a whole, therefore, it seemed clear that the null hypothesis should be rejected and that all three types of measurement seemed definitely and positively correlated. In general, the JR and physiological scores produced higher correlations than the PR.

When the data for female subjects were computed, it was found that almost all correlation values were smaller than for males and more were negative. Thus, the three most crucial values, corresponding to those listed immediately above for males, were:

PR-12 *vs* JR-12	.20 (NS)
Phys-12 *vs* PR-12	—.23 (NS)
Phys-12 *vs* JR-12	.36 (VS)

The question of sex differences will be considered in more detail immediately below. At this point in the analysis, it appeared that the females showed the same over-all trends as were reported for males in terms of correlations among the three types of measurement, but to a much smaller degree.

and "12" will symbolize Session One and Two combined. Thus: PR-1, PR-2, PR-12. To show circulatory data, "P" will symbolize pulse; "S," systolic blood pressure; "D," diastolic blood pressure; and "Phys," a combination of pulse, systolic and diastolic. Thus, P-1 symbolizes pulse taken during first experimental session; Phys-12, combination of all three circulatory measures for both sessions.

TABLE II

CORRELATION COEFFICIENTS FOR 50 MALE SUBJECTS BETWEEN MEASUREMENTS OF VARIOUS ASPECTS OF STAGE FRIGHT: INTROSPECTIVE INVENTORY (PR), AVERAGED JUDGES' RATINGS (PR), FLUCTUATIONS IN PULSE (P), SYSTOLIC (S) AND DIASTOLIC (D) BLOOD PRESSURE, THE THREE CIRCULATORY MEASURES COMBINED (PHYS); FOR TWO EXPERIMENTAL SESSIONS, SEPARATELY AND COMBINED.

	Session 1	Session 2	Sessions 1 + 2
PR-JR	.50	.16	.46
P-PR	.30	.45	.26
S-PR	.03	—.12	.10
D-PR	.06	.16	.34
PHYS-PR	.20	.26	.33
P-JR	.42	.45	.46
S-JR	.32	.54	.58
D-JR	.28	.31	.39
PHYS-JR	.48	.64	.64

Turning to the third main portion of the problem, sex differences in the data were studied. Comparisons of corresponding means in Table I indicated extensive divergencies.

The data showed that in both sessions the JR means for women were higher than for men. The *t*-ratios were: for the first session, 4.24 (VS); and for the second, 3.0 (VS). According to this, the women speakers *appeared* to be more confident than the men.

In both sessions the PR means for males were higher than those for females: The *t*-ratios were: 2.08 (S) and 1.22 (NS). According to this, there was a tendency, at least in Session One, for men speakers to rate themselves as *feeling* more confident. This, of course, contradicted the JR differences. The situation might facetiously be put: either the women were more accomplished as dissemblers or the men as boasters.

Mean differences between sexes in the data for pulse and blood pressure were also notable. The difference between means for normal pulse rate was not significant; that for normal diastolic blood pressure was very significant; and the normal systolic difference with a *t*-ratio of 1.9635 was just under the 5%

level. In Session One, the mean pulse rate was 6.3 points faster for females than for males, a significant difference; however, in Session Two the female pulse mean was only insignificantly higher. In both sessions, the male systolic rates were very significantly higher than the female. Likewise, in both sessions male diastolic readings were higher than female; in Session One this difference was not significant while in Session Two it was very significant. Thus, it appeared that on the average, in response to the speaking situation, women's pulse rate rose more than the men's; while men's blood pressure rose more than the women's. As has previously been indicated, physiologists tend to regard pulse fluctuation as indicative of more immediate, ephemeral disturbances; and blood pressure fluctuation (especially diastolic) as indicative of slower, more profound disturbances.

In discussing Table II, it was pointed out that in a substantial majority of instances, female correlation values were significantly smaller than the male. In marked contrast to males, over half the female correlation coefficients were insignificant; and most female *r*'s between PR and the various physiological measures were negative in sign, although insignificantly large.

Considered in perspective, there were so many significant sex differences, that it clearly necessitated treating the male and female scores as separate sets of data. The possibility was suggested that quite different measuring techniques might have to be applied to the two sexes. In fact, the correlation coefficients for PR indicated the high probability that two radically different introspective inventories should be devised. There was even the possibility that the experience of stage fright might be fundamentally different for men and women.

IV. Conclusions

1. The normal pulse and blood pressure rates of over 90% of the subjects were measurably affected by the speaking situation.

2. The amounts of circulatory fluctuation were marked; mean amounts of fluctuation were: 18.1 for pulse, 10.1 for systolic blood pressure, and 7.3 for diastolic blood pressure.

3. The direction of fluctuations was predominately upward although in a substantial minority of instances the rates decreased.

4. A very significantly larger number of subjects showed greater pulse fluctuation immediately before speaking than immediately after speaking.

5. Corresponding differences in blood pressure fluctuation before and after speaking were not statistically significant.

6. Among the male subjects, the physiological scores, and the PR (introspective) scores, and the JR (judges' rating) scores were definitely and positively correlated; most crucial *r* values were: PR-JR, .46 (VS); physiological-PR, .33 (S); physiological-JR, .64 (VS).

7. Among the female subjects, although the correlation values showed approximately the same relative relationships among the three types of measurement as for males, the *r*'s were much smaller, often negative, often not significant; most crucial *r* values were: PR-JR, .20 (NS); physiological-PR, —.23 (NS); physiological-JR, .36 (VS).

8. In general, the JR and physiological scores produced higher correlations than the PR.

9. The data showed many significant sex differences: women's scores were higher (toward the "fear" end) for PR and pulse fluctuation; men's scores, for JR and blood pressure fluctuation.

CHANGES IN CONFIDENCE DURING A PERIOD OF SPEECH TRAINING: TRANSFER OF TRAINING AND COMPARISON OF IMPROVED AND NON-IMPROVED GROUPS ON THE BELL ADJUSTMENT INVENTORY

STANLEY F. PAULSON
University of Minnesota

THIS study is concerned with changes of confidence in a group of students during a period of speech training. Its purpose is (1) to determine whether gains in confidence reported in a speech class tend to carry over to a different speaking situation, and (2) to ascertain whether those who report most improvement in confidence differ from those who report least improvement in confidence with respect to mean scores on the Bell *Adjustment Inventory*.

Reports have been published of several investigations concerned with changes of personality resulting from speech training. Moore[1] found significant improvement in self-sufficiency, dominance, emotional stability, and a significant decrease in introversion, as measured by the Bernreuter Personality Inventory. Rose[2] found a significant increase in dominance and decrease in neurotic tendency as measured by the Bernreuter Personality Inventory. Barnett[3] found a significant change in the direction of emotional extroversion as measured by the Minnesota Thinking-Social-Emotional Inventory. Leyden[4] found significant improvement in social confidence as measured by Knower's Speech Attitude Scale and the Minnesota Inventory of Social Behavior. Lomas[5] found that stage fright tended to decrease progressively during the speech training period, but when students spoke under conditions different from those in the class room, stage fright increased again. Gilkinson[6] found that improvement in confidence in the class room was accompanied by improvement in scores on an Inventory of Social Behavior. A study by Moore[7] compared students who improved most in action, voice, composition, ideas and poise, with those who improved least in those respects. He found no significant differences between the two groups on the Speech Attitude Scale, the Bernreuter Personality Inventory, and the Freshman Placement Examination.

METHOD

1. Two hundred seventy-one students in the first quarter of Fundamentals of Speech at the University of Minnesota filled out Gilkinson's Personal Report on Confidence as a Speaker (hereafter designated PRCS) immediately after

[1] Moore, Glen, "Personality Changes Resulting from Training in Fundamentals of Speech," SM, II (1935), pp. 56-59.

[2] Rose, Forrest A., "Training in Speech and Changes in Personality," *QJS*, 26 (1940), pp. 193-196.

[3] Barnett, Wynett, "An Experimental Study in the Teaching of Voice, Diction, through Ear Training, Phonetic and Oral Reading Approaches," SM, XV (1948), No. 2, 142-153.

[4] Leyden, Ralph, "An Experimental Study of the Effects of Speech Training at the Secondary School Level," M.A. Thesis, University of Minnesota, 1941.

[5] Lomas, Charles, "A Study of Stage Fright as Measured by Student Reactions," M.A. Thesis, Northwestern University, 1934.

[6] Gilkinson, Howard, "Indexes of Change in Attitudes and Behavior among Students Enrolled in General Speech Courses," SM, VIII (1941), 23-33.

[7] Moore, Wilbur E., "Factors Related to Achievement and Improvement in Public Speaking," *QJS*, 29 (1943), 213-217.

having made their first speech in class. After ten weeks of speech training, they again filled out the PRCS after having made a speech. First and final PRCS scores were compared to see if a statistically significant change had occurred.

2. Another group of fifty-six students filled out the PRCS after their first speech in class and again after a speech in class ten weeks later. They then gave their next speech in front of a Fundamentals section other than their own, i.e., before strangers, and again filled out the PRCS. Scores for these three speeches were compared.

3. The Bell Adjustment scores of the students in the larger group who gained most confidence were compared with those of the students who gained the least.

PRCS was designed by Gilkinson[8] "to develop a method of securing reports from students on the emotions which they experience in speaking before their classmates." The third part of the PRCS was used in this research. This part presents one hundred statements, the first fifty reflecting varying degrees of fear and the last fifty reflecting varying degrees of confidence. The lowest possible score would be —50 indicating the least amount of confidence; the highest possible score would be +50 indicating the highest level of confidence.

The comparison of those who improved most in confidence with those who improved least, raised special problems. If one simply selects those who improve most on the basis of the greatest improvement in scores from initial test to retest, he will find that they are the individuals who had low scores in the initial test. Their initial low scores make it easier to improve. Again, if one selects those who improve least on the basis of the least change in scores from initial test to retest, he will find that they are individuals who had high scores on the initial test. A person with an initial score of —30 (indicating little confidence) has 80 points along the scale as possibilities of improvement, while a person with an initial score of +30 has only 20 points along the scale as possibilities of improvement. An increase of 20 points, then, is quite different in these two cases due to the wide variation in initial position.

Also, from our knowledge of regression toward the mean, we know that if an individual is at one of the extremes of a distribution in his initial score, his score will probably be closer to the mean in the retest. Yet in one case this would mean "improvement in confidence" and in the other "loss of confidence." What is desirable in this case, then, is a comparison of improvers and non-improvers whose scores are approximately the same at the beginning. To accomplish this, the distribution of initial scores was divided into five equal segments of fifteen points along the scale. Within each one of these segments the twenty-five percent who improved most were compared with the twenty-five percent who improved least or not at all. Thus the difference due to initial place on the scale was minimized. This method of comparison has not, to the author's knowledge, been used previously in studies of speech improvement, and one of the most significant results of this study would not have appeared except through the utilization of this method.

After the improvers and non-improvers were selected in this fashion, the mean Bell Adjustment scores of the two groups were compared.

[8] Gilkinson, Howard, "Social Fears as Reported by Students in College Speech Classes," SM, IX (1942), 141-160.

Results

A. *Changes of PRCS Scores After Ten Weeks Training*

Table I shows the mean scores on the PRCS before and after ten weeks of speech training. Significant changes occurred for both men and women, indicating a rise in confidence. Women's scores are lower than those of the men but they reveal almost the same amount of improvement.

TABLE I

Comparison of Mean Scores on PRCS before and after Ten Weeks Training

N	Sex	Initial Scores	Final Scores	r	Diff./σdiff.
163	Men	1.05	9.60	.58	6.33
108	Women	—3.83	3.47		3.65

B. *Changes of PRCS Scores After Speaking Before Strangers*

Table II shows the mean scores on the PRCS before and after ten weeks of training and after giving a speech before a strange audience ("transfer score"). It will be noted that the average scores for both men and women increase slightly, though not significantly, when they spoke to a strange audience.

TABLE II

Comparison of Mean Scores on PRCS after Ten Weeks Training and after Speaking in Front of a Strange Audience

N	Sex	A Initial Score	B Ten Weeks Score	C Transfer Score
31	Men	—1.67	7.32	8.93
28	Women	—6.21	3.25	3.82
N	Sex	Diff./σdiff. A-B	Diff./σdiff. B-C	Diff./σdiff. A-C
31	Men	2.06	.38	2.33
28	Women	2.85	.17	3.03

C. *Comparison of the Bell Adjustment Inventory Scores of Improvers and Non-Improvers*

Table III provides a comparison of the improvers and non-improvers on the four Bell Adjustment Inventory scales: Home, Health, Social and Emotional. It will be recalled that the groups were selected from five segments of the range of PRCS scores in the initial test. The number of subjects in each segment is indicated, the range of that segment is given, and the mean Bell scores are stated for each segment as well as for the total groups. High scores on the Bell Inventory indicate poor adjustment and low scores indicate good adjustment. The critical ratios at the bottom of the table indicate the significance of the differences between the total groups on the four scores of the Inventory.

The table shows only slight differences in mean scores on Home, Health, and Emotional Adjustment. The greatest difference between the two groups occurs in the Social Adjustment scores. The improvers have the lower (more satisfactory) scores in each of the five segments of the PRCS range. The critical ratio of the difference between the two groups is 3.57 indicating statistical sig-

TABLE III

COMPARISON OF IMPROVERS AND NON-IMPROVERS AS TO MEAN SCORES ON THE BELL ADJUSTMENT INVENTORY

N	Range on PRCS	Home	Health	Social	Emotional
	Improvers				
5	26 to 40	8.1	5.0	2.8	7.2
11	11 to 25	4.7	5.3	4.8	5.7
20	−4 to 10	5.8	5.3	4.9	5.7
17	−19 to −5	6.5	6.9	8.0	7.8
7	−34 to −20	7.2	7.0	10.1	13.8
Means		5.74	5.34	6.10	8.00
	Non-Improvers				
N	Range on PRCS	Home	Health	Social	Emotional
5	26 to 40	6.2	6.6	3.8	4.6
11	11 to 25	7.8	5.1	5.3	6.0
20	−4 to 10	6.3	5.5	11.4	8.6
17	−19 to −5	5.9	7.4	10.8	10.2
7	−34 to −20	6.4	6.8	22.0	11.0
Means		6.24	5.56	10.60	8.05
Diff./σdiff.		.43	.23	3.57	.08

nificance. The table also indicates a positive correlation between confidence as reported in PRCS and good social adjustment as reported in the Bell Inventory.

INTERPRETATIONS AND CONCLUSIONS

A. *The Effect of a Strange Audience on Confidence*

One might expect that confidence gained during a period of speech training would be lowered when the subject leaves his own class and speaks in front of an audience of strangers. It has been suggested by some teachers that the rise in confidence experienced in a speech class is due to the fact that the students become well acquainted with each other, get accustomed to talking in front of a familiar group, and therefore feel more at ease. This implies that when a student leaves his own class to speak in front of an unfamiliar audience, he will probably not retain the confidence he acquired in his training class. The results secured in this study, however, show no decrease in confidence scores after the students had spoken in front of an audience of individuals with whom they were not acquainted. Apparently some transfer of training occurred. Theoretically, transfer of training is likely to occur when the transfer situation is basically similar to the training situation. In this study the training and transfer situations were similar in most respects: size of audience, type of audience, class-room conditions, length of speech, purpose of speech. It is not surprising, therefore, that the students showed no decrease of confidence in the new situation. This study deals with only one situation element; there is a need for experiments on the influence of other elements such as the size of the audience and the prestige of its members.

B. *Comparison of Improvers and Non-Improvers*

The one difference between those who gained confidence and those who did not was in social adjustment scores. Apparently, at the beginning of the course the

former were in general better adjusted socially than were the latter. This is a reasonable outcome, i.e., it is reasonable to expect that an individual who has made satisfactory adjustments to other social situations will do likewise in the speech class, and that an individual who had difficulty in adjusting to social situations before entering a speech class will have similar difficulties there. The results of this study emphasize the important fact that some speech students do not improve in confidence, and indicate that the failure to improve is often associated with poor social adjustment. It is further indicated that PRCS and the Bell Adjustment Inventory can be of assistance to teachers in discovering which students need special assistance. There is need for further experiments to discover what type of training is of most help to those students who at the beginning of the training period have low confidence associated with poor general social adjustment.

od intended to hold the initial PRCS scores relatively constant. The first part of Table IV shows that the method accomplished this purpose, there being only .25 points difference in the initial PRCS scores of improvers and non-improvers. The same section of the table also shows what would have happened if initial scores had not been held constant, i.e., if raw gains alone had been used as the basis of selecting the two groups. In that case there would have been a difference in the initial position of the two groups on the scale of 28.70 points. The second part of the table demonstrates that if initial position had not been controlled, no significant difference on the Bell Social Adjustment scores would have appeared.

Why is there so little difference between the Social Adjustment scores of the improvers and non-improvers when initial positions are uncontrolled? Apparently the difference in this case is due to the operation in the data of directly

TABLE IV

A Comparison of Results Obtained from Two Methods of Comparing Improvers and Non-Improvers

	Initial PRCS Scores			
	Improvers		Non-Improvers	Difference
Initial Position Controlled	—.15		.10	.25
Initial Position Uncontrolled	—14.50		14.20	28.70
	Bell Inventory Scores			
	Home	Health	Social	Emotional
Initial Position Controlled				
Improvers	5.74	5.34	6.10	8.00
Non-Improvers	6.24	5.56	10.60	8.05
Difference	.50	.22	4.50	.05
Initial Position Uncontrolled				
Improvers	5.68	5.70	8.95	7.90
Non-Improvers	6.43	6.14	9.05	8.50
Difference	.75	.44	.10	.60

C. *A Practical Method of Controlled Comparison*

As previously indicated, improvers and non-improvers were selected by a meth-

contradictory trends. Those who improve in confidence tend to have better Social Adjustment scores, but when the initial position is uncontrolled, they

have relatively low PRCS scores and this factor tends to give them poorer Social Adjustment scores, there being a correlation between PRCS and Social Adjustment (Table III). These two tendencies cancel each other. Conversely, those that do not improve in confidence tend to have poorer Social Adjustment scores, but when initial position is not controlled they also tend to have relatively high confidence scores which gives them better Social Adjustment scores. Here again the two factors work against each other and obliterate the difference.

The significant difference between the Social Adjustment scores of improvers and non-improvers appears in the controlled comparison because the factor of initial position is held constant. In this way the factor of improvement in confidence can be isolated and its relation to social adjustment shown. Therefore, when improvers and non-improvers are compared in studies of speech training, it would seem wise to make certain that the two groups are equated for initial scores.

D. *Conclusions*

1. Both men and women showed significant increases in confidence during ten weeks of speech training.

2. After a period of ten weeks of speech training, the mean scores of both men and women showed that improved confidence tended to remain even though they spoke to an audience composed of strangers.

3. A comparison of those who improved most in confidence with those who improved least showed no significant differences in their scores on Home Adjustment, Health Adjustment, and Emotional Adjustment scales of the Bell Inventory.

4. A comparison of those who improved most in confidence with those who improved least showed a significant difference in scores on the Social Adjustment scale of the Bell Adjustment Inventory. The mean Social Adjustment scores of those who improved most in confidence were more satisfactory than the scores of those who improved least.

5. No differentiation between those who improved most and those who improved least, in terms of their scores on the Bell Adjustment Inventory, appeared when these groups were chosen without regard to their initial positions on the PRCS.

THE RELATION OF PSYCHOMETRIC FACTORS TO STAGE FRIGHT

GORDON M. LOW, M.S., *Speech Clinician, University of Minnesota*
BOYD V. SHEETS, Ph.D., *Joint Director, Speech and Hearing Correction Center, University of Utah*

IN 1948 a change in policy at the University of Utah made it requisite that every freshman student register for a course in the Fundamentals of speech as a part of his general-education requirements. This new ruling automatically doubled the number of students taking this basic course and, at the same time, created problems in the effective handling of the increased teaching load. Among these problems was one which might have been anticipated by staff members engaged in clinical speech correction, namely, that of providing adequate training facilities for students who experience unusual emotional difficulty in speaking before a group. Previously, with speech as an elective course, a number of such students avoided this feared situation simply by electing other courses to satisfy university basic education requirements. Now, with the adoption of the new speech course ruling these students found themselves competing with their more confident classmates in classroom speaking situations.

During the first two years of the new program a number of cases of unusual stage fright were reported. Apparently, the experience of facing a group of fellow students even in a relatively informal speaking situation proved sufficiently traumatic to evoke profound emotional disturbances. In one instance, a male student, during his third short talk before his speech class, became so upset that he fled from the class and was not able to bring himself to return for the remainder of the quarter. This and other somewhat less dramatic cases pointed up the need for further information concerning students with stage fright who, because of apparent emotional difficulties, make a poor adjustment to the public speaking situation. The study presented here represents one of several areas being explored in an effort to assist the student who has an unusual amount of stage fright in meeting his problem.

An impressive number of writers in the field, including Gilkinson,[1] Knower,[2, 3] Murray[4] and Chenowith[5] have reported studies exploring the relationship between personality and speech adjustment and ability. Generally speaking, these reports tend to demonstrate an intimate relationship among these factors, and point out the additional fact that the poor speaker is most often the one who has had very little experience in facing an audience. New psychometric instruments have appeared in student counseling use and older tests have been revised. The present study takes advantage of certain of these more

[1] Gilkinson, Howard, A Questionnaire Study of the Causes of Social Fears among College Speech Students. *Speech Monogr.* 1943, 10, 74-83.

[2] Gilkinson, Howard and Knower, F. H., Psychological Studies of Individual Differences among Students of Speech. WPA Report Project #5209-96, University of Minnesota, 1939.

[3] Knower, Franklin H. A Study of Speech Attitudes and Adjustments *Speech Monographs*, 1938, 5, 130-203.

[4] Murray, Elwood, A Study of Factors Contributing to the Maladjustment of the Speech Personality. *Speech Monogr.* 1936, 3, 95-108.

[5] Chenowith, Eugene C., The Adjustment of College Freshmen to the Speaking Situation. *Quart. J. Speech*, 1940, 9, 585-588.

recent instruments in approaching the problem of psychometric factors in stage fright.

The manner of selection of the two criterion groups for the present study is different from most of those used previously, as is also the criterion of judgment. In many of the earlier studies the selection of criterion groups was made solely on the basis of the student's own evaluation of himself, the evaluation of a single instructor, or the judgment of auditors as to speech effectiveness. In the present study the choice of groups was made on the basis of (1) the student's own rating of himself, (2) the evaluation of his instructor, and (3) the judgments of his fellow class members. All of these ratings were made over a relatively extended period of time, with emphasis on the student's stage fright rather than his speaking ability.

Another aspect of the present investigation which was thought to be of importance was the fact that the total school population was well represented by the fundamentals of speech classes at the University of Utah, since all freshman students are required to enroll in the class. In contrast to this situation, many universities require a speech fundamentals class only of students in certain colleges.

The purpose of the present study, then, was to determine the relationship of stage fright to psychometric factors such as those purportedly measured by personality inventories, intelligence tests and attitude scales. The authors desired to compare the performance of students having a high degree of stage fright with those having very little stage fright on the basis of new group psychometric tests of the pencil-and-paper type. For the purposes of this project "stage fright" was defined as the emotional disturbances of mental and physical behavior of the speaker as manifested by poor eye-contact, nervous hand movements, restless shifting of feet, awkward posture, quivering of body, timid voice, embarrassment and other physical and vocal cues emphatically perceived.

The subjects used in this study were students enrolled in the fundamentals of speech course (Speech 1) at the University of Utah during the winter quarter of the school year 1949-50. From over 400 students enrolled in the fundamentals course, 132 were found to qualify as members of the criterion groups "least stage fright" (L), and "most stage fright" (M). The factors which determined whether or not a subject belonged in either the (L) or (M) group were: A judgment by the student's instructor at the end of a quarter's speech training, a judgment made by fellow class members at the end of the quarter, and the student's own rating of himself. All judges were instructed to consider the subject's total performance for the term, rather than his most recent or earliest efforts. The instructor was asked to rate his class in rank order as to the extent of each student's stage fright, placing him in one of three categories: Least Stage Fright, Average Stage Fright, or Most Stage Fright.

Students in the various classes were asked to rate themselves and each of their classmates in terms of the extent of stage fright observed. A nine-point rating scale was suggested and the list of stage fright characteristics cited above was offered as a point of reference for the student judge.

As a further index of the student's own estimate of his confidence in a public speaking situation, each of the subjects completed a Speech Questionnaire, consisting of a twenty question rating scale designed to obtain the student's report of his own stage fright. This

questionnaire included such questions as "I enjoy speaking to a group: (1) Not at all, (2) very little, (3) a little, (4) very much;" and "In speaking before a group I find that clear thinking is (1) almost impossible (2) exceedingly difficult, (3) rather difficult, (4) rather easy, (5) very easy."

The final placement of subjects into the appropriate criterion groups, i.e., "least stage fright" (L) and "most stage fright" (M) was accomplished in three steps. The first step was to determine the membership of tentative stage fright group (M). Students were assigned to this group on the basis of the ratings of fellow class members. Twenty-five per cent of the total membership of beginning speech classes was included in group (M) in this initial selection. The second step was to reduce the size of the initial group (M) by retaining as members only those students who were rated by their instructors as belonging to the group with the most stage fright.

The third step involved the reduction of tentative group (M) further by retaining only those students who themselves professed unusual stage fright. Thus only students were included in the "most stage fright" group who were judged as having most stage fright by their classmates, by their instructors and by themselves. A similar procedure was followed in the selection of the "least stage fright" (L) group. Neither students nor instructors were aware of the nature of evaluations made by other participants in the individual judgments.

Of the 132 students found to fit into either the (L) or (M) categories, sixty-seven were in the (M) group—twenty-one women and forty-six men. Sixty-five qualified for the (L) group, of whom eighteen were women and forty-seven were men. Test scores were not consistently available for all subjects on all the psychometric measures being considered. For this reason, the two criterion groups varied somewhat in size for the different tests. However, none of these groups contained fewer than forty subjects.

Psychometric Tests

With the (L) and (M) groups selected, it remained to compare them on the basis of scores made on several psychometric tests. The results of the battery of tests used in the present study were obtained from the files of the University of Utah Guidance Center. These examinations are administered each Fall to all entering freshmen. Thus, in almost every case, students had taken the examinations before enrolling in Fundamentals of Speech.

A brief description of the measuring instruments used in this study is given below:

The Cooperative English Test. Scores obtainable from this test include (1) mechanics of expression, (2) effectiveness of expression, (3) vocabulary, (4) speed of comprehension, and (5) level of comprehension.

The Cooperative General Achievement Tests. This group of tests was developed to provide measures of general proficiency in the fields of social studies, natural sciences, and mathematics, as contrasted with mastery of the more specific content of separate courses.

American Council on Education Psychological Examination for College Freshmen. The purpose of this test is to appraise what has been called scholastic aptitude or general intelligence, with special reference to the requirements of most college curricula.

Lee-Thorpe Occupational Interest Inventory—Advanced Series. This inventory seeks to determine the occupational

interests, not the abilities nor skills, of the testee. Of the results obtainable from this test, one of the scores was designed to assist in determining whether or not the individual is interested in verbal activities which involve facility in the use of language.

The Minnesota Multiphasic Personality Inventory. This psychometric instrument was designed ultimately to provide, in a single test, scores on all the more important phases of personality. Although the scales are designated according to the abnormal manifestations of the symptomatic complex, e.g. hysteria, paranoia, etc., they have all been shown to have meaning within the normal range.

The Biographic Inventory. This instrument consists of eighty-one questions concerning the past and present relationship of the testee with his parents, his friends of both sexes, and his associates in supervisory capacities. Other questions are intended to reveal the student's attitude toward his school work, sex, religion and social customs. Norms have not yet been established for this questionnaire.

Speech Questionnaire. This inventory consists of thirty-five multiple-choice questions. Twenty of the questions were designed to obtain a report of the student's confidence in the speaking situation. The remaining fifteen questions seek to determine the nature and extent of the subject's speech experience as well as his attitudes toward certain aspects of the speech situation.

In comparing the performance of the (L) and (M) groups on the different tests, "t" ratios, or critical ratios were calculated according to the formulas presented by Garrett[6] in his discussion of sampling and reliability. Differences at the .01 level of confidence were considered to be significant. For the Cooperative English Test, the Cooperative General Achievement Tests, the Psychological Examination, the Occupational Interest Inventory and the Minnesota Multiphasic Personality Inventory, comparisons were made by determining the reliability of the difference between the means of the test scores made by the two criterion groups. For the Biographic Inventory and the Speech Questionnaire, comparisons were made by determining the reliability of the difference between the percentages of the items chosen by the two groups in answering each question.

[6] Garrett, Henry E., *Statistics in Psychology and Education.* New York: Longmans, Green and Company, 1947.

Results

Based upon the treatment of the test results by the statistical methods outlined above, certain significant differences emerged between the group with most stage fright (M) and that with the least stage fright (L).

1. In the Cooperative Tests, significant differences appeared in sections on vocabulary (Critical Ratio 3.06), speech of comprehension (CR 2.86), total reading (CR 3.11), total English (CR 2.90) and social studies (CR 2.79). The poorer scores made by the (M) group would seem to indicate a general lack of ability on the part of the stage-frightened student to manipulate verbal concepts, as compared to the (L) group.

2. On the Psychological Examination statistically significant differences between the (L) and (M) groups occurred on the linguistic score of the test (CR 3.28). This measure indicates that students with most stage fright have less linguistic ability than do students with little stage fright.

3. The Occupational Interest Inventory reflected significant differences in

the following areas: Arts (CR 2.79), Verbal (CR 3.54) and Level of Interest (CR 3.12). These differences may be interpreted to mean that the student with the most stage fright apparently is not as interested in activities which involve self-expression, in verbal activities and in work involving judgment and the supervision of others, as is the student with the least stage fright.

4. When the scores on the Minnesota Multiphasic Personality Inventory were compared for the (L) and (M) groups, none of the differences were found to be significant at the .01 level of confidence. However, tendencies toward differences were indicated on three of the scales, namely the K (CR 2.48), the D (Depression) (CR 1.99) and the Pt (Psychasthenia) (CR 2.01). The K scale difference is significant at the .02 level of confidence, whereas the other two are significant at the .05 level of confidence. The differences indicate more of a lack of confidence and more of a tendency to worry on the part of the stage-frightened student. K score differences indicate that the stage-frightened student has a greater desire to make a poor score on the test than has the student with little stage fright.

5. When the results of the Biographic Inventory were considered it was found that on fifteen of the eighty-one questions there were significant differences in the responses chosen by the (L) group as against the (M) group. Differences in those questions pertinent to the study showed that students belonging to the group with the most stage fright are not as effective in their social relationships as students with the least amount. The (M) group indicated that their social adjustment before coming to college was generally poor. More men in the (M) group reported that they had not participated in school organizations nor taken part in extra-curricular activities than did men students in the (L) group. Both men and women in the (M) group reported that they had not been leaders of the organizations in which they were members, nor had they been leaders in their childhood.

The results of the Biographic Inventory also reflected that in their relationship with other people, students of the (M) group indicated that they were shy and rather withdrawn until they became better acquainted. Likewise their association with members of the opposite sex was limited, difficult and trying. A significant number of the stage-frightened students had not had their first date until they were more than fifteen years of age. This same group claimed to have more difficulty in expressing their feelings than did students with little stage fright.

6. The last of the test measures to yield significant differences between the (L) and (M) groups of students was the Speech Questionnaire. Of the fifteen questions designed to discover attitudes toward speech and the extent of speech experience of the student, thirteen showed significant differences between the two criterion groups.

The most important difference between the (L) and (M) groups was the lack of experience in public speaking of the (M) group. The critical ratio for this difference was the highest found in the entire study (CR 8.34). Students of the (L) group reported that they had more speech classes in high school than those of the (M) group, and that they had also appeared in more plays. A significant number of men of the (L) group had appeared before audiences in public speaking situations almost continuously for two years while serving as missionaries for the Church of Jesus Christ of Latter-Day Saints.

Members of the (M) group felt that what they had to say to an audience was not worth saying or not very interesting, and that the audience reaction to their speeches was unfriendly and cold or showed only mild interest. They considered their speech to be jerky, halting or slightly defective. If given their choice, students in the (M) category would participate in sports rather than write for publications, assume committee membership for school functions, sing or play a musical instrument or engage in speaking activities. Further, they would omit English composition from their college training.

Conclusions

In a study of the performance of students in beginning speech classes on certain psychometric tests, differences were noted between the criterion groups "most stage fright" and "least stage fright." The psychometric tests used in the study were: The American Council on Education Cooperatice English Tests; Cooperative General Achievement Tests, and Psychological Examination for College Freshmen; the Lee-Thorpe Occupational Interest Inventory, Advanced Series; the Minnesota Multiphasic Personality Inventory; a biographic inventory; and a speech questionnaire.

Test results indicated that students who have the most stage fright as compared with students who have the least stage fright:

1. Have not engaged in as much platform speaking activity.
2. Have not participated as much in extra-curricular and social activities.
3. Have difficulty in making an adequate social adjustment.
4. Have less linguistic ability.
5. Have less interest in activities which involve self-expression in verbal activities and in work involving judgment and the supervision of others.

No significant differences were found between students with the most stage fright and students with the least stage fright in:

1. General intelligence.
2. Quantitative reasoning ability.
3. The more important phases of personality.
4. Their interest in the fields of science, mechanics, nature and business.

Of greatest significance among the findings of this study was the difference in the amount of speaking experience reflected in the backgrounds of students with the most stage fright and those with the least amount. These findings are in agreement with those of Knower, Chenowith and of their studies of maladjusted and adjusted speakers. In making practical application of this apparent difference between the two groups of students, the obvious solution in reducing stage fright would seem to be placing fearful students in situations where they would get the requisite speaking experience. Although this approach would undoubtedly have merit in many instances, the results of this present study seem to point to the need for a broader interpretation of the findings. There were many indications, for example, that students with unusual stage fright did not lack for opportunities to obtain speaking experience. They came from essentially the same environment as did students with the least amount of stage fright, however, had consistently avoided speaking situations. This fact might be interpreted to mean that the lack of speaking experience was only symptomatic of a more deep-seated personality problem suggesting the need for a clinical approach to the more severe cases, either to precede or parallel the public speaking class.

RELATIONSHIPS BETWEEN VOICE VARIABLES AND SPEECH INTELLIGIBILITY IN HIGH LEVEL NOISE*

G. L. DRAEGERT**
Purdue University

THE transmission of information in a rapid and accurate manner is essential to the success of many organizations, especially those engaged in military type operations. Telephone and voice radio transmissions are often employed because of facility and rapidity of communications, yet costly inaccuracies often occur. Reduction or elimination of voice communication errors has been the subject of considerable research in the past decade. While much improvement has been made in communication systems, difficulties do occur which result in costly errors.

Elements of the communication situation where difficulties may commonly occur are: 1) message, 2) talker, 3) electrical system, and 4) listener. Analyses of the necessary tasks of each element have been performed at one or more of the following laboratories: a) Bell Telephone Laboratories; b) Psycho-Acoustic Laboratory, Harvard University; c) Electro-Acoustic Laboratory, Harvard University; d) Acoustic Laboratory, School of Aviation, Medicine and Research, N.A.S., Pensacola, Florida; e) Voice Communication Laboratory, formerly at Waco Army Field, Waco, Texas; and f) Voice Science Laboratory, Purdue University. Methods for training talkers have also been subjected to experimental techniques for evaluation.

In independent researches Abrams[1] and Moser[2] reported the intelligibility of language elements. Sounds, words, and combinations were recommended on the basis of experimental evidence. The present 'phonetic alphabet' used by the services resulted in part from such recommendations.

Voice characteristics of the talker and methods of training talkers have been reported by the laboratories listed above. Moore[3] reported that . . . loudness of voice is an important factor in voice communication and he cited evidence showing greater intelligibility accompanying speech signals of greater intensity. Curtis,[4] Abrams,[5] Miller,[6] Harris,[7] and Lightfoot and Morrill[8] reported im-

*This research was carried out under contract with the Office of Naval Research, Special Devices Center, Human Engineering Division, as Contract N6ori-104, T.O. II, Project NR-782-003, of which this is Technical Report No. SDC 104-2-25. The research report also constituted a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Purdue University. The research was directed by Dr. M. D. Steer and was conducted by the writer with the assistance of other members of the research project.

**Now affiliated with South Dakota College of Agriculture and Mechanical Arts.

[1] Abrams, M. H., "The Use of Words and Sentences in Testing the Intelligibility of Telephone Talkers," *Speech in Noise: A Study of the Factors Determining Its Intelligibility*, OSRD No. 4023, IC-69, 1944.

[2] Moser, Henry M., "Intelligibility Related to Routinized Messages," SM, XIII (1946), 47-49.

[3] Moore, Paul, "Intelligibility Related to Loudness," SM, XIII (1946), 13-18.

[4] Curtis, James F., "Intelligibility Related to Microphone Position, SM, XIII (1946), 8-12.

[5] Abrams, M. H., et al., "The Effect of Microphone Position on the Intelligibility of Speech in Noise," *Speech in Noise: A Study of the Factors Determining Its Intelligibility*, OSRD No. 4023, IC-54, 1943.

[6] Miller, G. A., et al., "Speech in Noise: A Study of the Attributes Determining its Intelligibility," *Speech in Noise: A Study of the Factors Determining Its Intelligibility*, OSRD No. 4023, IC-81, 1944.

[7] Harris, Jane S., "A Comparison of Rorschach and Minnesota Multiphasic Personality Inventory Test Results with Differences in Speaking Performance under Normal and Stress Conditions," MS. Thesis, Purdue University, 1948.

[8] Lightfoot, Charles and Morrill, Scott N., "Loudness of Speaking: the Effect of the Intensity of Sidetone upon the Intensity of the

proved intelligibility accompanied increased speech signal intensity. Mason[9] concluded that at high signal levels, increased speech intensity is advantageous only if the speech-to-noise ratio is improved. Where the speech-to-noise ratio was maintained at a constant value, a *decrease* in intelligibility accompanied an *increase* in signal intensity. Abrams[10] and Miller[11] found that intensity variability did not bear a close relationship to intelligibility and Harris[12] noted that increased intensity variability accompanied decreased intensity and lowered intelligibility for trained talkers placed under stress conditions.

Other researchers have been concerned with relationships between time factors and intelligibility. Harris[13] and Kelly[14] reported greater syllable duration to accompany higher intelligibility. Hanley and Steer[15] reported that uninstructed talkers reacted to increased noise barriers in a communication situation by increasing syllable duration and by decreasing speaking rate even without instruction to do so. Miller[16] found a similar tendency toward slower speech to accompany louder speech. Rate in words per minute, another aspect of time was investigated at the Waco Voice Communication Laboratory and in a report to OSRD[17] no clear superiority was indicated for any rate between 90 and 210 words per minute. Lightfoot and Black[18] reported that listeners respond to spoken messages with oral rates that are functions of stimulus rates.

Pitch is another voice variable investigated in connection with speech in high level noise. Miller[19] reported that among speakers controlling pitch, superior intelligibility scores were obtained by those talkers using higher pitches; but when intensity and other variables were held constant by statistical means there was a slight negative correlation between pitch and intelligibility. Brackett[20] noted that when intensity was controlled by the talker, a low correlation existed between pitch and intelligibility. He also reported a non-significant advantage in favor of uninstructed pitch when loudness was controlled by the talker. When intensity was controlled, Abrams[21] found no significant intelligibility difference between male and female voices, though known differences in habitual pitch level exist between the sexes. In spite of the often heard advice to 'pitch your voice up,' there seems to be no evidence from the various controlled experimental investigations to support the informal instruction.

Voice quality and its relation to intelligibility was investigated by Miller.[22] He found that quality described in terms of energy in various frequency bands or

Speaker," SDC 411-1-4, Acoustic Lab., Sch. Aviat. Med. and Res., NAS, Pensacola, Fla., 1948.

[9] Mason, Harry M., "The Effect of Very Loud Speech Signals upon Intelligibility," SM, XIII (1946), 19-23.

[10] Abrams, M. H., "Factors Related to the Intelligibility of Talkers in Noise," *Speech in Noise: A Study of the Factors Determining Its Intelligibility,* OSRD No. 4023, IC-60-1944.

[11] *Op. cit.*

[12] *Op. cit.*

[13] *Op. cit.*

[14] Kelly, J. C., "Effect of Training on Speech Intelligibility through Synthetic Noise Barriers," Ph.D. Dissertation, Purdue University, 1948.

[15] Hanley, T. D. and Steer, M. D., "Effect of Level of Distracting Noise upon Speaking Rate, Duration, and Intensity," *Journal of Speech and Hearing Disorders,* 14 (1949), 363-368.

[16] *Op. cit.*

[17] "Final Report in Summary of Work on Voice Communication," OSRD Report No. 5568, 1945, PB No. 12051, U. S. Dept. Commerce, 1946.

[18] Lightfoot, Charles and Black, John W., "Rates of Speaking: Reports II-III (Responses to Heard Stimuli)," SOC 411-1-3, Acoustic Lab., Sch. Aviat. Med. and Res., NAS., Pensacola, Fla., 1948.

[19] *Op. cit.*

[20] Bracket, I. P., "Intelligibility Related to Pitch," SM, XIII (1946), 24-31.

[21] Abrams, M. H., "Factors Related to the Intelligibility of Talkers in Noise," *Speech in Noise: A Study of the Factors Determining Its Intelligibility,* OSRD No. 4023, IC-60, 1944.

[22] *Op. cit.*

irregularity of spectra was not closely related to intelligibility. Correlations were not significantly different from zero.

In addition to the voice variables investigated experimentally, training programs reported by Kelly,[23] Shoup,[24] and Steer and Hadley,[25] recommended loud speech, slow speech, regular emphasis, 'effective' pitch levels, distinct articulation, and effective rhythms and phrasing.

In an attempt to find relationships, either positive or negative, when larger experimental populations were used, or when larger speech samples were used for measurements of voice variable, or when talkers were uninstructed, the investigation reported here was planned. The rationale used as a basis for the investigation was that if uninstructed speech of talkers exhibits certain characteristics which are closely related to intelligibility, then training to improve performance in those characteristics should be investigated for economy and effectiveness. Only those voice characteristics which were measurable with the instruments available were investigated with respect to intelligibility.

The speech sample used was a 24-word multiple-choice intelligibility test list accompanied by a 74 word prose passage. The subjects were 88 male undergraduate university students who were selected as the tails—upper and lower extremes—from a population of 555 men students. Criteria for selection were combined intelligibility scores and judgments of intelligibility based on impromptu speaking rated by 6-9 circuit mates.

After the 88 subjects had been selected, they were recorded reading the test passages, a phonetically treated paragraph and a test list of 24 words, over the electrical inter-communication system. A carbon microphone (type T-38), an amplified-noise generator (Device 8-I), and earphones (ANB-H-1A) comprised the system. High level noise was mixed with the side-tone in the talker's headset as he read the word list and passage. The speech signal from the carbon microphone was recorded by means of a Presto 8DG recorder with associated amplifiers, and the recorded samples were played through good quality playbacks into the desired instruments for measurements of the voice variables. The recorded test lists were also played to panels of listeners to determine the criterion intelligibility score to accompany the voice variable measures under consideration.

The variables and methods for determining them were:

1. Criterion intelligibility—obtained by playing the recorded test lists for the 88 subjects to panels of 10 listeners. As the records were presented to the listening panels, high-level noise was mixed with the side-tone in the headsets to provide a 'communication' barrier comparable to that experienced in the initial intelligibility survey.

2. Initial intelligibility score—determined by percent correct responses as the talker read a VCL test list[26] into a carbon microphone and Device 8-I interphone trainer. The initial score was obtained before selection of extremes when the individual was a part of the population of 555 students.

3. Initial rating score—obtained in the same experimental session as 2 above. The individual spoke in an impromptu manner for a period of approximately

[23] *Op. cit.*

[24] Shoup, Foster C., "Training Procedures," SM, XIII (1946), 59-63.

[25] Steer, M. D., and Hadley, J. M., "The Speech Intelligibility Program in Naval Aviation: Historical Summary," *QJS*, 32 (1946), 217-227.

[26] Haagen, C. Hess, "Intelligibility Measurement: Twenty-four Word Multiple-Choice Intelligibility Test," OSRD Report No. 5567, 1945.

30 seconds on a subject such as 'How to hitch-hike.' Circuit mates rated the speaker on a 1-5 scale of intelligibility ranging from *unintelligible* to *wholly intelligible.* The individual score was the combined judgment by his 6-9 circuit mates.

4. Mean syllable duration in seconds —obtained by playing recorded test lists into the Purdue Speech Sound Timer.[27] Those intervals during which speech energy exceeded a low threshold were accumulated to provide a measure of speech time. The mean syllable duration measure was calculated by dividing the accumulated speech time by the number of dictionary syllables in the speech sample.

5. Mean syllable intensity in db above an arbitrary reference—obtained by playing the recorded test list through an amplifier into a Sound Apparatus Company High Speed Power Level Recorder (model HPL-E). The graphic intensity traces were measured and the syllable peak magnitudes were tabulated. The mean peak values in decibels, as taken from the graphic records, were adjusted to compensate for initial attenuation in making the test records. The mean peak value in db was the measure used.

6. Intensity variability in db—determined by the standard deviation of the distribution of syllable peak magnitudes tabulated in 5 above.

7. Mean fundamental pitch—obtained from phonellograms of approximately ⅜ of the test words spoken. Phonograph records of the selected words were played into a Vibrograph,[28] an instrument to provide phonellograms. The mean period for each speech sample, converted to frequency by virtue of the reciprocal relationship between period and frequency, constituted the vocal frequency (pitch) measure used.

8. Harmonic structure factor—obtained by playing the recorded test list through a filter set (Western Electric RA-358) into the HPL-E graphic recorder. On the first playing, the filter was adjusted for *800 cps low pass.* The HPL-E graphic records were treated in the same manner as those for *4* above. Mean intensity for the second playing was subtracted from mean intensity for the first playing for the voice quality score. As measures of intensity in various frequency bands have been utilized to obtain indices of voice quality, the technique here described may be said to provide a gross means for determining the portion of harmonic spectrum in which sound energy is concentrated.

9. Rate in words per minute—for prose passages read and recorded immediately prior to the test list.

10. Proportion of speech time—obtained by determining *speech time* for the prose passage as in *4* above and dividing by elapsed time for the task.

11. Mean syllable duration in seconds —obtained by dividing *speech time* found for *10* above by the number of dictionary syllables recorded.

12. Mean syllable intensity in db above and arbitrary reference—determined for the prose passage in the same manner as *5* above.

13. Intensity variability in db—determined for the prose passage in the same manner as *6* above.

14. Sustained phonation in seconds—an independent measure of the talker's ability to sustain an *ah* sound judged acceptable by two clinicians. The average of three trials was used as a score.

15. Vital capacity in liters—obtained

[27] Tyler, H. J., and Draegert, G. L., "Purdue Speech Sound Timer," Voice Science Laboratory, Purdue University, N6ori-104, T.O. II, SDC-104-2-7, Lafayette, Indiana, 1948.

[28] Tiffin, Joseph and Steer, M. D., "The Vibrograph, a Combination Apparatus for the Speech Laboratory," *QJS,* 25 (1939).

from the average of three trials with a wet spirometer.

TABLE I
MEANS AND STANDARD DEVIATIONS FOR INTELLIGIBILITY TESTS AND VOICE VARIABLES FOR 88 SPEAKERS

Variable	M	S.D.
1. Criterion intelligibility scores	48.5%	18.3
2. Initial intelligibility scores	58.2	17.7
3. Initial rating	3.06	1.24
Word List		
4. Syllable duration in seconds	.256	.053
5. Mean syllable intensity in db (above arbitrary reference)	40.8	6.42
6. Intensity variability in db	7.58	1.36
7. Mean wave period on phonellogram in cm. (directly related to fundamental pitch)	.926[1]	.128
8. Harmonic structure indicated by difference between mean low frequency intensity and mean high frequency intensity in db	—.18	1.4
Prose Reading		
9. Rate in words per minute	158	24.7
10. Proportion of speech time to total reading time	.545	.096
11. Syllable duration in seconds	.157	.031
12. Mean syllable intensity in db (above arbitrary reference)	40.5	7.91
13. Intensity variability in db	7.10	1.07
14. Sustained phonation in seconds	21.8	5.72
15. Vital capacity in liters	4.50	.778

[1] Corresponds to a frequency of 184.3 cps.

The results are indicated in Tables I, II, and III. Distribution data for each of the 15 variables investigated in this study are shown in Table I. More complete descriptions of the variables appear in the preceding paragraphs. The values appearing in this table were derived from measures which had adequate internal consistency, but these values should not be compared with results of other studies without considering differences in situations, subjects, and equipment.

The 88 subjects were separated on the basis of the criterion intelligibility scores obtained from the recorded test lists. There were 44 above 50 per cent intelligibility and 44 below 50 per cent. The superior group of speakers was then compared to the inferior group with respect to all the other measured variables. The numbered variables shown in Table II are the same as those indicated in Table I and in the text.

Correlation coefficients were also computed between variables in all possible paired combinations. Tetrachoric correlation coefficients were first computed, using a formula which does not place

TABLE II
COMPARISON BETWEEN SPEAKERS OF SUPERIOR AND INFERIOR INTELLIGIBILITY READING WORD LISTS AND A PROSE PASSAGE

Variable*	Superior (n = 44) Mean	S.D.	Inferior (n = 44) Mean	S.D.	Diff.	t
1. Criterion intelligibility	63.3	8.17	33.3	11.6	30.3	93.3[1]
2. Initial intelligibility	66.9	14.3	49.5	16.3	17.4	5.29[1]
3. Initial rating	3.67	1.16	2.38	1.87	1.29	5.42[1]
4. Syllable duration	.285	.050	.227	.039	.058	6.02[1]
5. Mean syllable intensity	44.4	4.99	37.2	5.33	7.20	6.33[1]
6. Intensity variability	7.78	1.29	7.35	1.41	0.43	1.46
7. Phonellogram wave period	.909[2]	.141	.943[3]	.111	0.034	1.24
8. Harmonic structure factor	—.068	1.54	—.286	1.21	.218	.74
9. Reading rate	153	28.0	163.0	19.9	10.00	1.81
10. Speech time proportion	.585	.081	.504	.091	.081	4.30[1]
11. Syllable duration	.174	.028	.141	.025	.033	4.78[1]
12. Reading syllable intensity	48.1	5.22	38.6	5.02	9.5	8.48[1]
13. Reading intensity variability	6.14	1.11	6.05	.976	0.09	.37
14. Sustained phonation time	22.1	4.77	20.9	6.32	1.2	1.00
15. Vital capacity	4.58	.717	4.43	.792	0.15	.93

[1] Significant at the 1 per cent level of confidence.
[2] Corresponds to 181.1 cps.
[3] Corresponds to 181.3 cps.
*For fuller explanation of numbered variables, see Table I or text.

restrictions on the positions of dichotomic lines. Where a tetrachoric r of a magnitude greater than .50 was obtained a product-moment r was computed. Low value r's were left uncorrected. Extremely high value r's were corrected for hetrogeneity in those instances where the variance of the homogeneous population could be inferred from known data. These obtained coefficients of correlation are shown in Table III.

usually is placed upon intensity, rate and pitch. "Best guess" training procedures occasionally take into account variability in some of those attributes as well. In this connection it is of some interest to examine some of the results presented in Table III. As previously stated, no statistically significant relationship between intelligibility and mean fundamental pitch was found. However, a moderately high correlation

TABLE III

COEFFICIENTS OF CORRELATION BETWEEN VARIABLES
(All coefficients are Tetrachoric unless Otherwise Noted.)

Variable*	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.52[1]	.55[1]	.56[1]	.76[1]	.33	.07	.12	—.10	.35	.92[2]	.80[1]	.22	.16	.18
2		.92[1]	.48[1]	.46[1]	.05	—.02	—.40	—.19	.33	.41[1]	.40[1]	—.07	—.03	.10
3			.39[1]	.44[1]	.05	—.02	—.40	—.19	.33	.38[1]	.45[1]	—.07	—.03	.10
4				.40	.36	.11	—.13	—.23	.46[1]	.77[1]	.43	—.14	.24	.14
5					—.02	—.06	.05	.14	.34[1]	.44[1]	.92[1]	.13	.20	.06
6						—.16	.23	—.25	.04	.29[1]	.14	.33	.01	.10
7							—.29	.11	—.08	.00	—.19	—.05	.11	.10
8								.13	.01	.12	.19	.27	.02	.21
9									—.24	—.45[1]	.14	—.08	—.10	—.15
10										.69[1]	.53[1]	.16	.22	.10
11											.61[1]	.20	.23	.25
12												—.02	.09	.00
13													.25	.08
14														.36
15														

[1] Product-moment r's, uncorrected for hetrogeneity.
[2] Product-moment r, corrected for hetrogeneity.
*For explanation of numbered variables, see Table I or text.

Noteworthy among the correlation coefficients in Table III are those for intelligibility related to syllable duration and related to pitch. The criterion intelligibility test (variable No. 1) was mostly correlated with syllable duration for prose reading (variable No. 11). This correlation, corrected for hetrogeneity, was .92. Mean fundamental pitch for the test list (variable No. 7) was correlated .07 with intelligibility. The standard error of this tetrachoric coefficient of correlation was .18, indicating that the correlation was not significantly different from zero.

In training programs designed to improve voice communications, emphasis

coefficient of .76 is to be noted for intelligibility and intensity for the test list (variable No. 5). Rate for prose reading (variable No. 9), on the other hand, is seen to be related to intelligibility by a coefficient of —.10. Intensity variability for the test list (variable No. 6) and for prose reading (variable No. 13) are seen to be related by low coefficients (.33 and .22 respectively) to the intelligibility variable.

Based on the results of this investigation of voice variables for talkers using military type communication equipment in the presence of high level noise certain conclusions seem warranted, if the limitations imposed by the subjects, the

instrumentation and the situation are kept in mind. They are:

1. Voice variables most closely related to speaker intelligibility in high level noise are vocal intensity and syllable duration. This moderately high relationship is to be noted when the vocal intensity and syllable duration data are obtained from the criterion intelligibility test recordings, and also when they are obtained from a recorded prose reading passage.

2. Intensity variability, mean fundamental pitch, and the harmonic structure factor investigated (the predominance of high frequency or low frequency energy) are low value concomitants considered with respect to intelligibility.

3. Measures of the ability to sustain phonation and vital capacity are not closely related to any of the other variables investigated.

Results obtained from this investigation suggest that training for a high degree of syllable intensity accompanying prolonged syllables may be most effective in training for improved intelligibility. If talkers using a communication system in high level noise achieve increased syllable intensity and duration they probably will have improved their chances of being understood correctly.

THE PHONETIC ASPECT OF JOSHUA STEELE'S SYSTEM OF PROSODY*

JOHN B. NEWMAN
Queens College

IN *Prosodia Rationalis; or, an Essay towards establishing the Melody and Measure of Speech, to be expressed and perpetuated by Peculiar Symbols,* which was published in London in 1775,[1] Joshua Steele developed a system of prosody which was based upon the nature and process of speech. Though like all other prosodic systems in that it described the way in which written language is uttered,[2] Steele's system was distinct (and, as he claimed, rational) in that it was derived from the living, spoken language rather than from some abstract historical system of metrics. Thus, Steele described prosodic structures on the basis of oral rather than silent interpretations.[3] Meaning could be communicated only as a result of the production, transmission, and reception of the factors inherent in the language process (i.e., speech), and writing, which was simply a means of making it permanent, should record all rather than only selected parts of that process.

Steele attempted, therefore, by means of his "peculiar symbols," to indicate and to clarify those ekphonetic factors[4] of the language process not provided for by the grammar, syntax, style, sentence structure, and punctuation of the text. In order to do this, he adapted musical notation for speech purposes. Although at first he made use of the five line musical staff, he changed the system of musical notes, which indicated "the leaping tones" of the voice in song, to curved lines in order to indicate "the sliding tones" of the voice in speech.[5] In this way, he meant to score writing like a musical orchestration by indicating in-

*This paper is based upon research done for a doctoral dissertation, entitled "Joshua Steele: Prosody in Speech Education," which was completed in 1950 at New York University under the sponsorship of Professor Dorothy I. Mulgrave.

[1] *A Bibliography of Speech Education,* by Lester Thonssen and Elizabeth Fatherson (New York, 1939), lists the title of the 1775 edition as "An essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols" (p. 494) and the title of the 1779 edition as "Prosodia Rationalis; or an essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols. 2nd ed. enl." (p. 228). This distinction in the titles of the two editions is also made by John Nichols, who was the printer for both editions (*Literary Anecdotes of the Eighteenth Century* [London, 1813], 3:209 n), and by William Dickson, who edited Steele's letters and papers (*Mitigation of Slavery* [London, 1814], p. xxiii). A copy of the 1775 edition, bearing the title given, is in the Columbia University Library. It contains xvii, 193 pages. Two copies of the 1779 edition, bearing the title given, are in the New York Public Library. They each contain xvii, 243 pages. The fifty pages of "Additions" in the second edition add little to the theory set forth in the duplicate 193 pages of the first edition. (Cf. T. S. Omond, *English Metrists* [Oxford, 1921], p. 87).

Since there is in the Princeton University Library a copy of a 1775 edition, otherwise identical with the Columbia copy bearing the title of the New York Public copies of the 1779 edition, the work will hereinafter be referred to simply as *PR.* References to the first 193 pages will be found in either edition, while references to pages 194-243 will be found only in the second, "enlarged" edition.

[2] Enid Hamer, *The Metres of English Poetry* (London, 1930), p. 93.

[3] In this reference, note W. M. Parrish's remark on a scansion by George Saintsbury: "It seems, as Bliss Perry says, to please no one but the author, and, we might add, to be intelligible to no one but the author." (*Reading Aloud,* 2nd ed. rev. [New York, 1941], p. 292).

[4] Adapted from Willi Apel's term *ekphonetic notation,* which he defines in part as a system of "signs calling for special [vocal] inflections used to bring out grammatical peculiarities, such as questions, exclamations, affirmations, etc." (*Harvard Dictionary of Music* [Cambridge, Mass., 1945], p. 235).

[5] *PR,* p. 4-8 *et in passim.*

tonation, tempo, emphasis, prolongations, loudness, and pauses. Coming toward the end of the eighteenth century, such a method of transcribing speech would have had a considerable effect on the development of elocution.[6]

Steele was a prosodist, however—not an elocutionist. Though he scored the pitch, stress, duration, loudness, and pause heard in the oral reading of prose and verse, his purpose was the same as that of the traditional prosodists who scanned verse alone. It was Steele's intention to describe the structure of the language process rather than, like the elocutionists, to prescribe specific patterns of utterance. By means of his quasi-musical notation, he had shown the *prosodics*[7] of speech to consist of its melody (intonation) and measure (rhythm), and so he accounted for the ekphonetic factors of the language process.

Steele was quick to realize, however, that the phonetic factors of the language process were also inadequately indicated, much less clarified, by means of the orthography and word structure of written communication alone. At a time when orthoepists were struggling to indicate pronunciation by means of systems that did little more than ameliorate the confusions of the orthography, Steele desiderated a truly phonetic method of notation.[8] But he merely postulated phonetic transcription. It was up to his more famous contemporaries, notably Thomas Sheridan and John Walker, to more or less fulfill the lexicographical need for a system of re-spelling for pronunciation with their cruder diacritical systems of numbered vowels.[9]

Steele was not an orthoepist, however; and *PR* was not a book on spelling and pronunciation.[10] As a prosodist, Steele was concerned primarily with speech rhythms rather than pronunciation. His consideration of the quality of sounds in pronunciation, however, rose out of his concern for the quantity of sounds in speech rhythms.[11] Steele's account of the vowels and diphthongs,[12] brief and sketchy though it was, was the entire extent of his phonetic investigation. A subordinate part of "the immediate subject" of *PR*, it served Steele's musical[13] and rhetorical analysis of speech.[14]

The resemblances between the symbols for the vowels and diphthongs that Steele provided and those of our present day International Phonetic Alphabet have already been noted.[15] Steele represented each sound by a separate symbol and used that symbol to represent only the sound it stood for, regardless of the spelling. In order to do this, he borrowed letters from the Greek alphabet and made use of both the upper and lower cases of the letters of the Roman.

[6] See John B. Newman, *Joshua Steele: Prosody in Speech Education* (New York University Ph.D. dissertation, 1950), pp. 176-234 *et in passim.*

[7] Cf. G. H. Monrad-Krohn's use of the term *prosodics* in "Dysprosody or Altered 'Melody of Language,'" *Brain—A Journal of Neurology*, Vol. 70, Part IV, p. 405.

[8] *PR*, pp. x-xi.

[9] Thomas Sheridan, *A Complete Dictionary of the English Language*, 2nd ed. (London, 1789), p. ix ff. (Originally published in 1780.) John Walker, *A Critical Pronouncing Dictionary* (London, 1791), p. 72.

[10] Cf. C. M. Wise, "Benjamin Franklin as a Phonetician," *Speech Monographs*, 15 (1948): 112.

[11] For examples of the interrelation of the quantity and quality of speech sounds, see J. S. Kenyon and T. A. Knott, *A Pronouncing Dictionary of American English* (Springfield, Mass., 1944), p. xxv; C. K. Thomas, *An Introduction to the Phonetics of American English* (New York, 1947), p. 109, 111; *Webster's New International*, p. xxxv ff.

[12] *PR*, pp. viii-xiii.

[13] Steele thought of speech as a separate genus of music. [*PR*, p. xvi ff].

[14] Steele was probably influenced by the ancient Greek concept that phonetics "was simply a handmaid in the service of music and rhetoric." [W. Rhys Roberts, *Dionysius of Halicarnassus on Literary Composition* (London, 1910), p. 43].

[15] M. M. Robb, *Oral Interpretation of Literature* (New York, 1941), p. 62.

This was far in advance of the practice of the times. Contemporary orthoepists used only the twenty-six letters of the Roman alphabet and were naively bound by orthography in their attempts to indicate pronunciation. Consequently they were at times forced to duplicate their notations. John Walker, for instance, represented the vowel in "fall, wall, water" and the vowel in "nor, for" as a^3 and o^3, respectively, despite his own admission that the sound the two symbols stood for was "perfectly equivalent."[16]

Steele's plea for the establishment of "a rational orthography" is also remarkable for its prophecy:

> . . . what very great advantages might arise to the lingual and literary commerce of the world, by a set of learned men sitting down, under some respectable authority, to reform the alphabet, so as to make it contain distinct elementary marks for expressing all the lingual sounds of the European languages at least; in doing which, the difficulty would be infinitely over-balanced by the great and general utility.[17]

But Steele's synthesis was not equal to his analysis.[18] His brief phonetic account of the sounds of speech is full of errors of omission as well as errors of commission. In reference to the consonants, for instance, he merely "throws out the hint" that "some very useful alterations and additions might be made among [them] towards attaining a rational orthography."[19] Further consideration is entirely omitted. From a prosodic point of view, however, such an omission is entirely conceivable. The consonants have very little, if any, quantity, sonority, or syllabification-ability, and so would afford slight interest to a prosodist.

Consideration of the vowels and diphthongs, however, is basic to prosody, and so Steele devoted the bulk of his interest to it. He even went so far as to establish definitions. A vowel, for instance, was defined as:

> . . . a simple sound capable of being continued invariably the same for a long time (for example, as long as the breath lasts), without any change of the organs; that is, without any movement of the throat, tongue, lips, or jaws.[20]

A. J. Ellis points out that this definition misses the essential characteristic of a vowel, namely, that it is "properly a musical tone with a definite quality or timbre."[21] Yet he admits the great success of Steele's endeavor "to write down speech," despite the fact that it did not enter his scheme to represent the quality of sound.[22] It would seem that here Steele's synthesis succeeded in spite of his analysis.

Steel observed that for all the vowel sounds:

> . . . the principal nations of Europe [i.e., those employing the Roman alphabet] use only five characters (for the *y* has, with us, no sound distinct from *i*), and this defect throws the orthography and pronunciation of the whole into uncertainty and confusion.[23]

Steele narrowed his considerations, however, by limiting the number of vowels. "There are, in nature, neither more, nor less, than seven vowel sounds, besides diphthongs."[24] The number must refer to the vowels that should be represented alphabetically; for Steele was "of opinion that every one of the seven has both a longer and shorter sound."[25]

The seven "natural" vowel sounds are "marked and explained to sound in English" as follows:

[16] Walker, *Pronouncing Dictionary*, p. 22, 11, 72.
[17] *PR*, p. xiii.
[18] Omond, *English Metrists*, pp. 92-3.
[19] *PR*, p. xiii.
[20] *Ibid.*, pp. viii-ix.
[21] *On Early English Pronunciation*, 5v. (London, 1869-89), 1:51. Cf. J. S. Kenyon, *American Pronunciation*, 8th ed. rev. (Ann Arbor, Mich., 1940), p. 57.
[22] *Ibid.*, 4:1055.
[23] *PR*, p. viii.
[24] *Ibid.*
[25] *Ibid.*, p. xii.

ɑ = all, small, or, for, knock, lock, *o*ccur.
a = man, can, cat, rat.
e = may, day, take, nation.
i = *e*vil, keen, it, be, *ini*quity.
o = *o*pen, *o*nly, bro*o*ke, h*o*le.
ω = fool, two, rule, tool, do.
u = *su*perfl*u*ous, tune, *su*preme, cred*u*lity (very rare in English).[26]

Their longer and shorter components are marked as follows:
ɑ is long in ɑ*ll*, and short in *lock* and *oc*[*cur*] (*l*ack and ɑc[*cur*]).
A is long in *arm*, and short in *cat*.
E is long in *may* and *make*, and short in *nation*.
I is long in *be*, and short in *it*.
O is longer in h*o*le than in *o*pen; long in corr*o*de, short in corr*o*sive.
Ω is long in *fool*, short (by comparison) in f*oo*lish.
U is long in *tune* and *plus*, and short in s*u*per and *du*.[27]

This table from *PR* shows how Steele used separate symbols to represent what were as far as he was concerned separate sounds. Here the longer and shorter components of the same vowels, which Steele thought of as separate sounds, are distinguished in their representation by the use of upper and lower cases of the same symbol, respectively, whether the symbol be a letter from the Greek or the Roman alphabet. Thus, the longer component of the vowel, as in the word *all*, is represented by an upper case symbol, viz., [ɑll], whereas the shorter component of the same vowel, as in the word *lock* and in such a word as *occur*, is represented by a lower case symbol, viz., (l*a*ck and ɑc[cur]).

Steele then goes on to explain that:

> The French, the Scotch, and the Welsh, use all these vowel sounds in their common pronunciation; but the English seldom or never sound the U in the French tone (which I have set down as the last in the foregoing list, and which, I believe, was the sound of the Greek [upsilon]) except in the more refined tone of the court, where it begins to obtain in a few words.[28]

John Odell, an early proponent of the Steele system of prosody, who also gave examples of "the longer and shorter sounds" of the seven vowels, did not recognize "the U in the French tone" as an English vowel at all.[29] Ellis later vouched for its occurrence in English as late as 1775; but cited Steele as evidence.[30] At any rate, it was undoubtedly no longer heard by the beginning of the nineteenth century, when Odell was writing, and so he substituted for it in his list the vowel [ʌ], which Steele had erroneously described as a diphthong.[31] According to Odell, [ʌ] is a short vowel with no longer counterpart.

In reference to Steele's long and short forms of a,[32] which are also recognized by Odell, it is interesting to note that in the alphabet of the *Association Phonétique Internationale* (IPA), [ɔ] as in *hall* and [ɒ] as in *hot* are symbolized as longer and shorter forms of the same sound.[33]

Steele's relation of the vowel in *arm* with that in *cat* as longer and shorter

[26] *Ibid.*, p. x.
[27] *Ibid.*, p. xii.
[28] *Ibid.*
[29] *An Essay on the Elements, Accents, and Prosody of the English Language* (London, 1806), p. 7. Odell's list of the vowels reads as follows:

1.	aw	ɑ	ɑ̄	Sol, Saul [ɒ, ɔ]
2.	ah	a	ā	ban, balm [æ, ɑ]
3.	a	e	ē	pen, pane [e, e:] or [ɛ, e:]
4.	e	i	ī	tin, teen [ɪ, i]
5.	o	o	ō	no, known [o, o:]
6.	oo	w	w̄	pull, pool [ʊ, u]
7.	ut	u		cull, come, cover [ʌ].

[30] *Early English Pronunciation*, 3:980 n. Ellis identifies the sound as [Y].
[31] *V. inf.*
[32] Underlined symbols are those given in the source cited. Bracketed symbols are my interpretations in IPA.
[33] O. Jespersen, *Growth and Structure of the English Language*, 4th ed. (New York, 1923), p. 253.

forms of the same sound is matched by Odell's pairing of *balm* and *ban*. Kenyon points out that:

> . . . standard speech up to about 1775 had no [ɑ] sound in . . . such words as *father, calm, hardly,* and such words as *ask, staff, jaunt, aunt, half,* etc., which now have [ɑ] in some types of English. Sheridan's pronouncing dictionary (London, 1780) shows no [ɑ]. Benjamin Franklin's phonetic transcriptions in 1768 have none. His pronunciation of *father, hardly* was [fæðr, hærdlɪ]. Noah Webster in 1789 has [æ:] in *aunt, jaunt, sauce.* E. Hale in 1799 has [æ] in *aft, balm, carve, gaunt,* etc.[34]

The vowels in *arm, balm, ban,* and *cat,* then, were pronounced the same in the standard speech of the time. Steele's consideration of the vowels in *arm* and *cat* as longer and shorter components of the same sound is purely quantitative or prosodic in view of his pronunciation. The fact that the phonetic quality of the vowel in *arm* changed since his time should not confuse our appraisal of Steele's analysis, which must be considered from the point of view of mid-eighteenth century pronunciation rather than from that of a later period (i.e., our own). This point of view should be maintained in considering the sounds that follow as well.

John Walker wondered "what light can . . . be thrown on this subject by [Joshua Steele] who, notwithstanding the infinitesimal distinctions he makes between similar sounds, says, . . . that the *a* in *may* is long, and the same letter in *nation* short."[35] This criticism does not mean anything, however, because Walker and Steele were thinking of the manifestations of different elements in the same instances: to Walker, "long *a*" meant [eɪ], "short *a*" [æ]; to Steele, "long" and "short" referred to duration. Qualitatively, the letter *a* in *may* and in *nation* is the same.[36] Quantitatively, however, the spoken vowels represented by the same letter in these two instances are distinct; the duration of the *a*-sound in *may* is longer than that in *nation*. Steele distinguished these two sounds on the basis of their quantity rather than their quality. In fact, the distinction between all of "the longer and shorter sounds" in Steele's list was meant to be quantitative rather than qualitative. That there happens in some cases to be a qualitative distinction that is now recognized as phonetic is simply the result of the immediacy of the interrelation of quantity and quality in the speaking process.

Odell's distinction between "long ē" in *pane* and "short e̱" in *pen* may have been meant to be qualitative or phonetic rather than quantitative or prosodic. If "long ē" was [e:] and "short e̱" [e], then the distinction may be said to have been prosodic. On the other hand, if "long ē" was [e:] and "short e̱" [ɛ], then the distinction must be said to have been phonetic. The difference in the quantity of the vowels in these two instances is so great as compared to their difference in quality[37] that Odell's distinction of them may be thought of as having been prosodic or quantitative in line with the others in the list.

Steele's distinction between "long I̱" in *be* and "short i̱" in *it* (Odell's *teen* and *tin*) is also meant to be one of syllabic quantity rather than phonetic quality, despite the qualitative difference.

The distinction in length between the pre-sonant *o* in *hole* and the pre-surd *o*

[34] Kenyon, *American Pronunciation,* pp. 172-3.

[35] *A Key to the Classical Pronunciation of Greek, Latin, and Scripture Proper Names,* annexed to "A Critical Pronouncing Dictionary and Expositor of the English Language," (New York, 1819), p. 91 n. (Originally published in London in 1798.)

[36] Walker marked both *a* in his *Dictionary.*

[37] Daniel Jones, who marks the vowel in *pen* [e], says: "Some authors write the sound with the sign [ɛ], and there is much to be said in favour of this mode of representing it." [*An English Pronouncing Dictionary,* rev. ed. (New York, 1931), p. xxii].

in *open* is still another example of the interrelation and interdependence of prosody and phonetics. Though quality is indeed a factor, Steele's distinction, again, is not meant to be qualitative at all. The vowel in both cases is recognized as qualitatively the same but quantitatively greater in *hole* than in *open*. Odell must have had a similar intention in distinguishing the *o* in *known* as long and the *o* in *no* as short.

More evident of Steele's prosodic or quantitative, rather than phonetic or qualitative, approach is his indication of greater length in the monosyllabic *fool* as compared to the same element in the disyllabic *foolish*. In both cases, the vowels are exactly the same in quality, but distinctly different in quantity. Odell's distinction between "long w̄" in *pool* and "short w̲" in *pull* brings the problem again into the realm of phonetics, for the vowels in these instances are different in quality.[38] A prosodist, listening for rhythm and metre, might sanction such a near rhyme as *pool-pull*; a phonetician, listening primarily for vowel quality, would have to distinguish the two sounds on a phonetic basis.

Steele's identification of the *u* in *tune* with the unrounded front vowel, as in the French *plus*, presents a confusing picture. No longer is it as simple as explaining the phonetics of prosody or the prosody of phonetics. Linguistic history and geography must now be considered, and in the darkness of Steele's "imperfect" French pronunciation. He anticipated the criticism of his French pronunciation[39] when he explained that:

> In order to ascertain the tones of the seven vocal sounds, I have ventured to add a few French words in the exemplification; in the pronunciation of which, I hope, I am not mistaken. If I had not thought it absolutely necessary, I would not have presumed to meddle with any living language but my own; the candid reader will therefore forgive and correct my errors, if I have made any in this place, by substituting such other French syllables as will answer the end proposed.[40]

Ellis' belief that "the U in the French tone" persisted in English up through the latter part of the eighteenth century lends credence to Steele's identification of it. But this would indicate that all of the French and English pairs in the list were meant to be identical. Were that actually the case, Steele's French must have, at best, sounded very English.

A glance at the French exemplifications in Steele's list will bear this out. *a* in "all, small, or, for, knock, lock, *oc*[cur]," for instance, is said to be heard "in French as the words en, grande;" *a* in "man, can, cat, rat" is said to be heard in the French "P*a*ris, *ha*bit, p*a*rdon;" *e* in "may, day, take, nation" is heard in "ses, et;" *i* in "*e*vil, keen, it, be, *in*iqu*i*ty" is heard in "Par*is*, hab*it*, ris, dit, it;" *o* in "*o*pen, *o*nly, br*o*ke, h*o*le" is heard in"s*o*ldat, c*ô*tes, *o*ffrir;" w̄ in "fool, two, rule, tool, do" is heard in "ou, vous, jour, jal*oux*;" and *u* in "s*u*perfl*u*ous, tune, s*u*preme, cred*u*lity" is heard in "du, plus, une."[41]

Despite their phonemic difference, however, the value of French exemplifications "to ascertain the tones" of the vowels of English seems to have also been recognized by John Walker, in his "Table of Simple and Diphthongal Vowels."[42] Walker did not make the same mistakes that Steele did, but his pronunciation of French seems to have been equally "broken in accent." Solely on the basis of who came first, it can be as-

[38] A possible explanation is Odell's mistaken orthographic analogy of *pool* with *wool*.

[39] Cf. Ellis, 4:1057.

[40] *PR*, p. x.

[41] *Ibid.*

[42] *Critical Pronouncing Dictionary*, p. 72. There are no French exemplifications in the 1st ed. of 1791. The earliest edition in which they appear that I have seen is the 1st American (Philadelphia, 1803), p. cxxxii.

sumed that Walker borrowed the idea of French exemplifications from Steele's table in the preface to *PR*.[43]

Steele's treatment of the diphthongs was neither as thorough nor as fortunate as his account of the vowels. In the first place, Steele was confused in his attempt to distinguish between an orthographic analysis of written language and an acoustic analysis of spoken language. Secondly, Steele was confused because he overstepped the bounds of prosody, willy-nilly, into the realm of phonetics. Prosodically, a vocalic sound, whether "simple" or "diphthongal," has quantity, sonority, and syllabification-ability. To distinguish between vowels and diphthongs on a purely prosodic basis necessitates minute quantitative analysis; and, as has been pointed out, Steele's synthesis was not equal to it.

The traditional concept of a diphthong was perhaps best expressed by John Walker:

> A diphthong is a double vowel, or the union or mixture of two vowels pronounced . . . so closely . . . together, so as . . . to form only the time of . . . one syllable: as the Latin *ae* or æ, *oe* or œ, the Greek ει, the English *ai*, *au*, &c.[44]

Aware that digraphs and diphthongs were indiscriminately called simply "diphthongs," Steele noted that:

> The letters and sounds, which in modern languages pass under the names of diphthongs, are of such different kinds, that they cannot be known by any definition I have seen: for, according to my sense, the greatest part of them are not diphthongs.[45]

Steele decided, therefore, that a "proper" diphthong was a *speech* sound, that is, a blend of two *heard* sounds rather than two letters written for one sound.[46]

On the basis of this distinction, Steele defined a diphthong as a sound:

> . . . made by blending two *vowel sounds*, by a very quick pronunciation, . . . so intimately into *one*, that the ear shall hardly be able to distinguish more than one uniform sound; though, if produced for a longer time than usual, it will be found to continue in a sound different from that on which it began, or from its *diphthong sound*.[47]

But whereas a vowel can be "continued invariably the same" for as long as the breath lasts without any movement of the speech organs, a diphthong:

> . . . most commonly changes immediately from the first vowel sound, of which [it] is composed, by a small movement in some of the organs, to the sound of the vowel which makes the latter part of the said diphthong, the sound of the first vowel being heard only for one instant.[48]

Harmless as this definition and description may appear, it created virtual havoc in Steele's list of the diphthongs. Thinking in prosodic terms, "blending two vowel sounds, by a very quick pronunciation, . . . intimately into one," means linking the length of two vowels into the time of a single syllable,[49] rather than "gliding on from one sound on to another."[50] Thus, Steele identified the sound of [ʌ] as a diphthong! This "English sound of U, as in the words UGLY, UNDONE, BUT, and GUT," he explained:

> . . . is composed of the English sounds AU and OO; but they require to be pronounced so extremely short and close together, that, in the endeavour to prolong the sound [in order to distinguish its component parts as described above], the voice will be in a continual confused struggle between the two component sounds, without making either of them, or any other sound, distinct; so that the true English sound of this diphthong can never be expressed but by the aid of a short energetic aspiration something like a short cough, which makes it

[43] For a full discussion of Walker's "borrowings" from Steele, see Newman, pp. 176-90.
[44] *Critical Pronouncing Dictionary*, p. 25.
[45] *PR*, p. xi.
[46] *Ibid.* Cf. Kenyon, *American Pronunciation*, p. 203.
[47] *PR*, pp. ix, xi.
[48] *Ibid.*, p. ix.
[49] See Walker's definition of a diphthong *supra*.
[50] Ellis, 4:1057.

very difficult to our Southern neighbours in Europe.[51]

Here Steele "seems to confuse a diphthong, in which there is a real succession of vowel sounds and a connecting glide, with the attempt to pronounce two vowels simultaneously [i.e., quantitatively in the time of one syllable]."[52]

In describing the sound of [ɛ] as a diphthong, Steele proved his analysis to have been equally confused:

> The English sound of E in the words *met, let, men, get,* is a diphthong composed of the vocal sounds A [as in "man, can, cat, rat"] and E [as in "may, day, take, nation"] . . . and pronounced very short.[53]

So much for the errors of commission in Steele's account of the diphthongs. In describing "the English sound of I or Y," however, both of which letters, he was quick to point out, "are the marks of one and the same diphthong sound," he identified the first element as "the sound AU, [that] immediately changes to EE on which it continues and ends."[54] This analysis was probably the reason for his having "altogether omitted to notice *oy*, and hence [having] escaped falling under the necessity of distinguishing *by, boy,* for example."[55] Presuming Steele to have been an Irishman,[56] Ellis thinks he would have lengthened the second element in *by* while in *boy* he would have lengthened the first—in Ellis' notation: "(bAii, bAAi)."[57]

The remaining two "diphthong sounds in English" in Steele's list of five are iw in "you, use, new, due, few (a long sound);" and oω in "how, bough, sow, hour, gown, town (this diphthong is sounded long, dwelling chiefly on the latter vowel)."[58]

Thus, Joshua Steele's analysis of the sounds of English speech was merely an aspect of his system of prosody. Though intonation and speech rhythms were his immediate subject, his description of the vowels and diphthongs as speech sounds, and his provision of symbols for their notation regardless of the spelling of the words in which they occurred, gave to his "rational prosody" a phonetic aspect as well. Though full of errors of omission as well as of commission, his account is remarkable for the perspicacity of its observations. Steele's view of the dynamics of the speech process as a manifold unity, and his projection of a tonetic as well as a phonetic method of notation, anticipated conclusions speech scientists were not to reach for a century after his time. His proposal, as far back as the latter part of the eighteenth century, for an international phonetic association was truly prophetic.

It did not enter Steele's scheme to represent the quality of sound. As a prosodist, he would be more immediately concerned with quantity. Yet he was hardly a proponent of quantity as the particular agency which, in Saintsbury's words, constitutes the difference of the value of syllables out of which rhythms and metre are made. The gist of Steele's system lay in the proposition that that difference in value was the result of the coeval influence of several separate and individual *speech* agencies, namely, pitch, stress, duration, volume, and pause. His phonetic analysis of the vowels and diphthongs hinges on the interrelation of the quantity and quality of speech sounds. In fact, his errors in phonetics, or the comprehension of the qualitative element of speech sounds, are the result of his prosodic method or

[51] *PR,* p. ix.
[52] Ellis, 4:1056. Cf. Kenyon, *American Pronunciation,* p. 204.
[53] *PR,* p. x.
[54] *Ibid.,* pp. ix-x.
[55] Ellis, 4:1056.
[56] As far as can be determined, Ireland was Steele's place of origin. See Newman, p. 22.
[57] Ellis, 4:1056.
[58] *PR,* p. xi.

quantitative approach to the analysis of the vowels and diphthongs.

Although Joshua Steele presented his work as a system of prosody, it approached the analysis of literary (i.e., written) structures on the basis of the physical requirements of their being read aloud. His work should, therefore, be considered primarily as a study of the process of speech. Else why should a prosodist be concerned with the "rationality" of orthography, a description of the vowels and diphthongs *per se,* and the establishment of an authority to reform the alphabet to provide for the expression of "all the lingual sounds of the European languages at least," much less the ekphonetic notation of such purely vocal factors as pitch, stress, duration, volume, and pause? It is possible that the identification of Steele's system with the mechanical method of oral interpretation has overshadowed his early contributions to the study of the nature of speech as a physical phenomenon. If such has been the case, the fact that his system of prosody had a phonetic aspect at all may help reorient widespread misconceptions of the contribution of Joshua Steele to the field of speech.

AN ORAL READING EVALUATION OF GOOD AND POOR SILENT READERS

WILLIAM B. McCOARD, *University of Southern California*
and
SAMUEL N. LeCOUNT, *Chico State College*

MYTH and misinformation becloud the relations between silent and oral reading. Practicalities have forced superintendents, principals, and teachers to proceed largely on "intuitive" basis. The Speech field generally has ignored the problem and left it to the Education Department. Some educators have suggested that the processes are quite different from each other; others have contended that training in one phase was good training for the other, too—that the processes are quite similar. Gates[1] says that oral reading is a more complex and difficult activity than silent reading—in oral reading one is required to do all that is demanded in silent reading and several things in addition. Because of the complexities of the problems involved, adequate research has been meager.

In order to make a study of the relationship of silent to oral reading, good measures of both types of reading are necessary. Fairly satisfactory progress has been made in the development of silent reading tests, enough so that the major problem of this study was to get adequate evaluations of oral reading ability. The fact that oral communication or speech can be evaluated only in terms of its effect upon each individual listener precludes standardized, objective, impersonal tests. Therefore, scientific studies have generally used the analysis of errors (substitutions, omissions, insertions, etc.) rather than broader measures of the effectiveness of the communication through oral reading. Some helpful studies have been reported by Traxler and Townsend,[2] Gaines,[3] Swanson,[4] Fairbanks,[5] Bond,[6] Carter,[7] Jones,[8] Mead,[9] Paul,[10] Pintner and Gilliland,[11] and Kelly.[12]

The present study was planned to investigate a broader approach to the problem. It attempted to find out whether any relationship could be shown between silent reading abifity as usually measured and ability in oral reading

1 Gates, A. I., *The Improvement of Reading*, 3rd ed. (Macmillan Co., 1947).

2 Traxler, Arthur E. and Townsend, Agatha, *Another Five Years of Research in Reading* (New York, 1946).

3 Gaines, Frances P., "Interrelations of Speech and Reading Disabilities," *Elem. School Journal*, XLI (April, 1941), 605-613.

4 Swanson, Donald E., "Common Elements in Silent and Oral Reading," *Psych. Monographs*, Vol. 48, No. 3, 36-60.

5 Fairbanks, Grant, "The Relation Between Eye Movements and Voice in the Oral Reading of Good and Poor Silent Readers," *Psych. Monographs*, XLVIII (1937), 78-107.

6 Bond, Guy L., "The Auditory and Speech Characteristics of Poor Readers," *Teachers College Contributions to Education*, No. 657 (New York: Teachers College, Columbia University, 1935).

7 Carter, Bertha, "A Study of the Oral Reading Disabilities of a Sixth, Seventh, and Eighth Grade Group Deficient in Silent Reading" (Master's Thesis, University of Iowa, Iowa City, Iowa, August, 1940).

8 Jones, Effie B., "A Comparison of Comprehension Results in Oral and Silent Reading," *Peabody Journal of Education*, IX (March, 1932), 292-96.

9 Mead, Cyrus D., "Results in Silent vs. Oral Reading," *Journal of Educ. Psych.*, VIII (June, 1917), 367-68.

10 Paul, Vera Alice, "The Relation of Oral Reading to Remedial Reading in Elementary and Secondary Schools," *QJS*, XXIX, No. 2, 217-22.

11 Pintner, Rudolph and Gilliland, A. R., "Oral and Silent Reading," *Journal of Educ. Psych.*, VII (April, 1916), 201-12.

12 Kelly, G. A., "Some Common Factors in Reading and Speech Disabilities," *Psych. Monographs*, XLIII (1932).

when appraised in terms of its broader, communicative effectiveness. Thus it was concerned with the relationships between silent reading comprehension abilities of selected college students and their abilities to project that comprehension in oral reading.

Subjects: The subjects were selected on the basis of total scores obtained in the Cooperative Reading Comprehension Test[13] which was administered to 658 new students at the University of California, College of Agriculture, at Davis, California. The highest fifty were chosen as the good reader group and the lowest fifty as the poor reader group. Compared with the national norms,[14] the mean score of the good reader group was in the 95th percentile and that of the poor reader group was in the 5th percentile.

Experimental Procedure: Each of the subjects was given a 230-word, simple narrative selection taken from the *Reader's Digest* entitled "The Spy Who Double-Crossed Hitler," which was read into the microphone of a Meissner disc recorder. To put each subject at ease, a quick demonstration was given to show how the machine worked. Uniform instructions were given to each of the 100 subjects as follows:

> You are to read this page aloud as meaningfully and as clearly as possible. If you come to a word with which you are not familiar, do not skip over it. You will not be timed as you read. Now read the first sentence for practice while I adjust the machine to your voice."

After the short practice reading, the cutting head was placed in position and as soon as it had made two or three revolutions on the disc, the subject was told to start reading. The recordings were made on 10-inch, acetate discs, at 33⅓ revolutions per minute.

All the subjects were recorded under similar conditions in an effort to lessen external variables. Room, position in room, microphone position, instructions, practice, recorder, records, and all other possible factors were kept as constant as possible. Each recording, then, represented the recorded response of a student to a standardized test situation.

Six speech experts were chosen to make evaluations of these oral readings. They averaged 7.3 years of experience in the teaching of speech at the college level. Two of the judges held the Ph.D. degree and four held Master's degrees in the field of speech.

Before listening to the recorded performances of the 100 subjects, the judges were given a preliminary practice period with fifteen recordings which were not part of the experimental group. The written instructions given the judges were as follows:

> The first fifteen records will serve as orientation and training records. They are not part of the group under study. They are a mixture of (a) five students who made high scores in the silent reading test but were just below the fifty highest selected for the study, (b) five students who made scores midway between the high and the low, and (c) five students whose scores were low but not in the bottom fifty selected for the study. As trained teachers of speech it can be assumed that you are acquainted with the responses likely to be made by college students; however this orientation period will further define and clarify (1) the subjects, (2) the materials, and (3) the oral responses of students who are somewhat comparable to the subject students in their silent reading scores. The addition of responses from some students with average silent reading scores is included as a further aid to orientation.

The rating scale used by the judges consisted of eleven items with a scale of from 1 to 7. Four represented the response of the average college student, 1 represented the poorer or poorest, and

[13] Cooperative English Test: Reading Comprehension, New York: Cooperative Test Service.

[14] American Council on Education, "College Norms for the Revised Series of the Cooperative English Tests" (Washington, D. C., 1941).

7 the better or best. The eleven items observed and rated were: general effectiveness, communication of ideas, communication of emotion, distinctness and general pronunciation level, pitch, variation of pitch, voice quality, variation in volume, rate, variation of rate, and phrasing. These items were chosen in an effort to secure as accurate judgments as possible and minimize effects of personal prejudice. It was thought, for example, that combining the responses to general items with responses to certain specific factors might give a more nearly accurate evaluation than would be obtained from a single judgment of a specific or general item.

The 100 recordings were played for the judges in a random order. The entire evaluation was completed in one period, broken only by a one-hour recess for lunch. Refreshments were served during the judging to help relieve fatigue.

A total oral reading score was obtained by summing up the ratings given by each judge for each item, then averaging the subjects' scores given by the six judges.[15]

Results: The six judges appeared to be in general agreement on their speech ratings (total score) of the subjects. The ratings of each judge were correlated with the ratings of each of the other judges. The *average intercorrelation* was found to be .69. The Spearman-Brown prophecy formula[16] was used with this average intercorrelation to ascertain what the reliability of the judgments would be for the six judges combined, and a predicted reliability of .93 was obtained. A reliability of .93 based on this type of rating scale of speech factors may be regarded as a high reliability.

Frequency distributions of the speech scores made by the good and poor silent reader-groups are shown in Table I.

TABLE I

FREQUENCY DISTRIBUTION OF GOOD AND POOR READERS ON SCORES OBTAINED IN SPEECH TEST

Speech Scores	Frequency	
	Good Readers	Poor Readers
15-19	1	5
20-24	4	5
25-29	4	13
30-34	9	17
35-39	7	6
40-44	10	1
45-49	9	2
50-54	3	1
55-59	2	
60-64	1	

The *mean* of the oral reading scores for the good silent readers was 40.8 (+10) while for the poor silent readers it was 31.0 (+7.5). The difference between the means was 9.8. The *t* ratio was 5.5, which was far greater than that required for significance at the 1 percent level of confidence. In other words, the good silent readers got reliably higher oral reading ratings than a corresponding group of poor readers.

An examination of Table I shows that the oral reading scores for the good and poor silent reader groups overlapped —with a possible range of 11 to 77 (66 points). The poor readers ranged from 17.5 to 52.5 (35 points) and the good readers ranged from 17.5 to 62.5 (45 points). However, five times as many poor silent readers received the lowest oral reading score as did good silent readers; and the highest oral score of the poor silent readers was 10 points

[15] In a comparable study (Wm. B. McCoard, "Speech Factors as Related to Teaching Efficiency," University of Wisconsin, unpublished Ph.D. thesis, 1941, pp. 145-148), it was determined through intercorrelations and composites that, in such a study, no score on any specific factor, or composite of some of the factors, could be regarded as more reliable than the method of using total speech scores. Factors of validity and reliability of scores of this type were also analyzed.

[16] Guilford, J. P., *Psychometric Methods* (New York: McGraw-Hill Book Co., 1935), 368n, 418-421.

lower than the highest score for the good silent readers.

The *mean* for the entire group of 100 subjects was 35.9. Further examination showed that 80 percent of the poor silent readers had oral reading scores below the mean, while only 30 percent of the good silent readers fell below the group mean on their oral reading scores. Reversing this, it was seen that 20 percent of the poor silent readers and 70 percent of the good silent readers had scores above the group mean for oral reading.

Conclusion: Other studies[17] have shown that good silent readers generally have fewer errors (substitutions, omissions, insertions, repetitions, etc.) when reading orally. The present study shows that when the oral reading of a college group is appraised in terms of the *broader aspects of speech* (general effectiveness, phrasing, variety of pitch, etc.), by a group of speech teachers, the extremes (good and poor) of a random distribution of *silent* readers show a significant difference in their *oral* reading abilities.

Incidental Observation: An incidental, yet interesting, point was illustrated through the use of the American Council on Education (ACE) Psychological Examination. The good and poor silent readers ranked respectively high and low in "academic intelligence." The average ACE score for the good reader group was in the 97th percentile of the national norms and that of the poor silent reader group ranked in the 17th percentile of national norms.[18] This further supports the concept that desirable traits and abilities tend to go together.[19] It might also be implied, on the basis of the relationships shown in the major study, that good speech habits may be included in such a group of positive factors; the chances are that a group showing high average intelligence and a superior silent reading ability will also show above the average ability in the broader communicative factors of oral reading. However, the correlations that have been reported by Gerber,[20] West and Larson,[21] Knower,[22] Gilkinson and Knower,[23] Eckert and Keyes,[24] and Dow and Papp,[25] between intelligence and student skills in speaking have been too low to have any predictive value. It is indicated, therefore, that further investigations should be made before final conclusions can be drawn.

[17] Swanson, *op. cit.;* Fairbanks, *op. cit.;* Anderson, Irving H. and Swanson, Donald E., "Common Factors in Eye Movements in Silent and Oral Reading," *Psych. Monographs,* XLVIII (1937), 61-69; LeCount, Samuel N., "A Study of the Oral Reading Abilities of Good and Poor Silent Readers with Emphasis on the Process of Word Recognition," Ph.D. Thesis (University of California, 1948).

[18] Thurstone, L. L., and Thurstone, T. G., "Psychological Examinations, 1941 Norms," (American Council on Education, Washington, D. C., 1942).

[19] Anastasi, Anne, *Differential Psychology* (Macmillan Co., 1937); Terman, L. M., et al., *Genetic Studies of Genius,* Vol. I: *Mental and Physical Traits of a Thousand Gifted Children* (Stanford, 1925).

[20] Gerber, John C., "Testing and Evaluation in the Skills of Communication," *College English* (April, 1948), 375-384.

[21] West, Robert E. and Larson, Helen, "Some Statistical Investigations in the Field of Speech," *Quarterly Journal of Speech Education* (November 1921) 7:379-80.

[22] Knower, Franklin, "Psychological Tests in Public Speaking," *Quarterly Journal of Speech* (April 1929) 15:211-217.

[23] Gilkinson, Howard and Knower, Franklin, "A Study of Standardized Personality Tests and Skill in Speech," *Journal of Educational Psych.* (1941) 31:161-175.

[24] Eckert, R. G. and Keyes, N., "Public Speaking as a Clue to Personality," *Journal of Applied Psychology* (1940), 24:144-153.

[25] Dow, Clyde W., and Papp, Stephen, "The Relation of Reading Ability and Language Ability to Speech Ability," *Speech Monographs* (1943).

AN EXPERIMENTAL STUDY OF THE EFFECTS OF SPEECH ORGANIZATION UPON ATTITUDES OF COLLEGE STUDENTS*

RAYMOND G. SMITH
Indiana University

INTRODUCTION

STUDIES of various factors affecting persuasion have appeared in the professional literature with ever increasing frequency since the announcement of the invention of a shift-of-opinion audience measuring instrument in 1922.[1] Using the experimental approach, the present study was directed toward determining the effect on the persuasive qualities of a speech of transposing its main parts. The effects of such modifications of the organizational pattern were then measured against the effects of the normal order speech to find out if speech organization, *per se*, is an important factor in persuasion.

Because it was possible to secure a great deal of additional information with a relatively small expenditure of time and effort, a number of questions in addition to the primary ones dealing with attitude change were asked of the experimental subjects. All such test items were selected on the basis of their possible relationship to social attitude, and included such factors as importance, knowledge, and interest in the topic. The study, then, was designed to answer, within the limitations of experimental technique, such questions as:

1. What is the effect on the persuasive qualities of a speech of modifying its organizational pattern?
2. Is there any relationship between social attitude change and any of the following factors:
   a. the importance with which the topic is regarded by the subject,
   b. the subject's interest in the topic,
   c. the subject's opinion of the extent of his knowledge about the topic,
   d. the subject's willingness to discuss the topic,
   e. liking the speech,
   f. liking the speaker,
   g. considering the speech to be organized, or
   h. considering the speech to be effective?
3. Are there relationships between any two of the above named factors?

It should be stated at the outset that this study deals ony with opinion; in all instances where it was possible to do so the questions were patterned after the Woodward shift-of-opinion ballot with

*Based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Wisconsin, 1950, and prepared under the direction of Andrew T. Weaver.

[1] Utterback, W. E. "Measuring the Reaction of an Audience to an Argumentative Speech," *QJS*, VIII (April, 1922), 181-3.

Collins, G. R. "The Relative Effectiveness of Condensed and Extended Motive Appeal," *QJS*, X (June, 1924), 221-30.

Woodward, H. S. "Measurement and Analysis of Audience Opinions," *QJS*, XIV (February, 1928), 94-111.

Millson, W. A. D. "A View of Research in Audience Reaction," *QJS*, XXIV (October, 1938), Part I, 464-83.

———. "Problems in Measuring Audience Reaction," *QJS*, XVIII (November, 1932), 621-37.

Wilkie, W. H. "An Experimental Comparison of the Speech, the Radio, and the Printed Page as Propaganda Devices," *Archives of Psychology*, XXV (1934), No. 169.

Knower, F. H. "A Study of the Effect of Oral Argument on Changes of Attitude," *Journal of Social Psychology*, VI (August, 1935), 315-47.

———. "A Study of the Effect of Printed Argument on Changes of Attitude," *Journal of Abnormal and Social Psychology*, XXX (January-March, 1936), 522-32.

———. "Some Incidence of Attitude Changes," *Journal of Applied Psychology*, XX (February, 1936), 114-27.

Lund, F. H. "The Psychology of Belief," *Journal of Abnormal and Social Psychology*, XX (1925) 68-81, 174-96.

three categories for initial and five for final expression of opinion.

The independent variable for the study was the various organizations of the speech as heard by the different experimental audiences. The dependent variable was the change in degree of audience response resulting from these respective speech modifications.

Subjects

Two separate experiments were run with different audiences. The subjects tested were members of the beginning public speaking course at Indiana University, none of whom had any previous college speech training. A total of 762 students was used; the majority were freshmen, but the sections included a sprinkling of sophomores, juniors, and seniors. The age range was from 17 to 40 years, but by far the largest number were from 18 to 20 years of age. There were no specially constituted sections, e.g., sections made up of selected students. Fifty-one students from three sections were used to procure a reliability check of the measuring instruments.[2] Three hundred sixty-five students heard the first experimental speech, 346 heard the second. Twenty-seven different sections constituted the audiences for Experiment I, twenty-four for Experiment II. At least three sections were combined in each instance to form each experimental group. After the groups had been equated on the basis of initial social attitude scores, there was a final total of 315 subjects constituting 9 groups of 35 persons each for the first experiment, and 252 subjects constituting seven groups of 36 persons each for the second experiment.

[2] The Pearson product-moment coefficient of correlation was .86, significant at the 1 per cent level.

Measuring Instruments

The Woodward shift-of-opinion type questions on the pre- and post-test forms used for Experiment I provided for registering opinion change for eight factors, seven of which it was assumed might be related to the key factor—social attitude shift. These eight questions as they appeared on the ballot were:

1. How important do you believe this subject is in your life?
2. In your opinion how great is your knowledge of this question?
3. What is your attitude toward other people's drinking?
4. Are you interested in this subject?
5. Would you like to participate in a discussion of this subject?
6. What is your attitude toward drinking for your personal use?
7. Do you believe that the moderate use of alcoholic beverages is harmful to the human body?
8. What is your attitude toward moderate social drinking?

For constructing data tables, and discussing results, key words were chosen to represent the above questions. These words, in order, were as follows:

1. *importance*
2. *knowledge*
3. *social attitude*
4. *interest*
5. *willingness to discuss*
6. *personal attitude*
7. *health*
8. *moderate social drinking*

The post-test form included four additional questions as follows:

1. How do you think you would like the speaker as a person?
2. How did you like the speech?
3. In your opinion how well was the speech organized?
4. In your opinion how effective was the speech?

Key words chosen to represent these questions respectively were:

1. *liking speaker*

2. *liking speech*
3. *believing speech organized*
4. *believing speech effective*

Questions six, seven, and eight of the pre-test ballot for Experiment I had to do specifically with the use of alcoholic beverages, hence these were dropped from the ballot of Experiment II. Otherwise the questions were retained with substantially the same wording.

The ballots were scored so that any shift, no matter how great, was given a point value of one. This was the manner of scoring used by Millson and checked by Monroe. Concerning the scoring of the ballot, Millson writes:

> Shift of opinion is treated as moving in one of two directions—toward the affirmative or negative. Every shift is given the value of one and scored for affirmative or negative.[3]

Monroe's conclusion concerning this method of scoring was:

> The shift-of-opinion ballot scored in the manner suggested by Millson, seems to be a reasonably valid measure of the shift-of-opinion produced in an audience by a speaker.[4]

In addition to being checked for retest reliability, the individual questions were checked with a Chi-square test for normalcy of distribution.

Experimental Speeches

Two separate experimental situations were set up, each using a ten-minute experimental speech. The topic for the first was *the effect of alcohol on the human organism*; it contained twelve minutes of carefully worded argument and appeal opposed to the practice of drinking alcoholic beverages. This speech consisted of an introduction, three main points, a conclusion, and transitions between the main points. Care was exercised to include in each of these sections material and only material of the nature generally agreed upon by rhetoricians as belonging within the section in question.

The second experimental speech was on the topic of *socialized medicine* and was adapted from the one used by Dr. Franklyn Haiman[5] in a series of experiments conducted at Northwestern University. For the present study this speech was divided into six main parts: attention, need, satisfaction, visualization, refutation, and action.

Each speech was transcribed on one side of a sixteen-inch broadcast quality acetate disc. In making the impressions, the divisions of the speech were slightly separated so that they could be precisely "spotted" when the recording was reproduced. The same speaker recorded each speech.

Physical Equipment

The physical equipment consisted of an RCA standard model recording-reproducing machine with light, speaker, and microphone connections for operating individually or simultaneously in four classrooms.

Temporal Sequence of Procedure

The experimental orders of the speech were as follows:

A. The two transitions were moved to the beginning of the speech.
B. The normal order speech.
C. The introduction was moved to the middle.
D. The random order speech.
E. The introduction and conclusion were reversed.
F. The transitions were moved to the end.
G. The conclusion was moved to the beginning.
H. The introduction was moved to the end.
I. The conclusion was moved to the middle.

The procedure consisted of distributing and collecting the pre-test forms of

[3] Millson, W. A. D. "A View of Research in Audience Reaction," *QJS*, XXIV (December, 1938), 6.

[4] Monroe, A. H. "The Reliability and Validity of the Shift-of-Opinion Ballot," *QJS*, XXIII (December, 1937), 585.

[5] Haiman, F. S. "An Experimental Study of the Effects of Ethos in Public Speaking," *SM*, XVI (1949), 190-202.

the ballot, playing the speech in the part order scheduled for the hour, and distributing and collecting the post-test forms.

Equating the Groups

Audiences were equated on the basis of initial attitude scores by randomly eliminating individuals from categories with observed frequencies of initial response higher than that expected theoretically for normalcy of distribution. Each experimental group was then checked against the normal order group by a Chi-square test for goodness of fit, and a T-test of similarity of variances, to assure equally constituted audiences.

The second experiment was conducted similarly to the first. The difference between the two experiments were first, that the speech topic was different, and second, that no transitions were designated, *per se,* in the second speech. The lack of transitions resulted in two less experimental variations for speech number two.

Methods of Control

All voice variables were perfectly controlled by the process of recording the speeches. The record playback was held constant by the use of mechanical controls on the recording and playback instrument. The audiences were controlled by randomizing the selection of the speech organizations and by using a minimum of three classes to make up each audience in each experimental variation. One operator made the recordings and playbacks. The four classrooms were similarly decorated, of approximately the same size and shape, and with the same placement of the permanently installed and identical wall speakers.

Results from Experiment I

Examination of Table I reveals a statistically significant change in mean scores for the normal order (criterion) group for but four of the eight questions asked. These questions concerned *social attitude, interest, health* and *moderate social drinking.* The gains for each of the nine experimental audiences were therefore calculated for each of these four significant factors. Table II lists the gains, mean gain, and standard error of the mean for each group for these four factors. The significance of the differences in mean gain between each experimental organization and the normal order group for each of the four significant factors were then calculated. Tables III, IV, V, and VI show this information. These tables indicate there was no significant difference between the mean gains of the normal order group and each of the experimental orders for any of the four factors checked, and it follows, therefore, that for this speech the sequence of the main parts made no real difference. Table VII lists in numerical order the critical ratios of the mean gains for the social attitude question of all nine experimental orders. It may be seen that statistically significant changes in mean gain with critical ratios above 3.0 occurred in three of the nine groups. The normal order speech

TABLE I

Experiment I

Gains on the Woodward Ballot for the Normal Order (Criterion) Group
(N = 35)

Question	Gain	Mean gain	σ_M	C.R.
1. Importance	6	.1714	.0945	1.8
2. Knowledge	—2	—.0572	.1951	.29
3. Social Attitude	—12	—.3429	.0897	3.82
4. Interest	31	.8857	.0785	11.3
5. Willingness to discuss	0	.0000	.0000	.00
6. Personal Attitude	—6	—.1714	.0856	2.0
7. Health	27	.7714	.0709	10.9
8. Moderate social drinking	—17	—.4857	.0936	5.2

TABLE II

Experiment I

GAINS FOR ALL GROUPS

(N = 35)

Order*	Social Attitude			Interest			Health			Moderate Social Drinking		
	Total gain	Mean gain	σ mean	Total gain	Mean gain	σ mean	Total gain	Mean gain	σ mean	Total gain	Mean gain	σ mean
A	—14	—.4000	.0921	32	.9143	.0473	24	.6857	.0785	—22	—.6826	.0755
B	—12	—.3429	.0898	31	.8857	.0538	27	.7714	.0709	—17	—.4857	.0936
C	—11	—.3143	.1051	31	.8857	.0673	24	.6857	.0883	—15	—.4286	.0929
D	—11	—.3143	.0971	28	.8000	.0788	26	.7429	.0840	—12	—.3429	.0852
E	—7	—.2000	.1054	31	.8857	.0673	21	.6000	.1006	—13	—.3714	.0997
F	—6	—.1714	.0856	29	.8286	.0754	20	.5714	.1085	—9	—.2571	.1022
G	—7	—.2000	.0676	27	.7714	.0716	22	.6826	.5057	—9	—.2571	.1018
H	—1	—.0286	.0851	31	.8857	.0673	20	.5714	.1033	—13	—.3714	.0812
I	—8	—.2286	.0911	29	.8286	.0637	24	.6857	.0881	—15	—.4286	.0837

*See order key, page 294.

yielded a greater shift of opinion than did any variation except that of moving the transitions to the beginning of the speech.

Pearson product-moment coefficients of correlation for the relationships among the gains of the various factors were determined as were the correlations between gain in social attitude and a. liking the speech, b. liking the speaker, c. believing the speech to be organized, and d. believing the speech to be effective. The conclusions of this paper will indicate that, for the most part, these correlations were not significant.

TABLE III

Experiment I

T-TEST OF THE SIGNIFICANCE OF THE DIFFERENCES IN MEAN GAINS IN SOCIAL ATTITUDE BETWEEN THE NORMAL ORDER (CRITERION) GROUP AND THE EXPERIMENTAL GROUPS

(N = 35)

Orders	Difference	Sigma of difference	C.R.
M_B-M_A	.0571	.13	.44
M_B-M_C	.0286	.14	.20
M_B-M_D	.0286	.13	.22
M_B-M_E	.1429	.14	1.02
M_B-M_F	.1715	.12	1.43
M_B-M_G	.1429	.11	1.29
M_B-M_H	.3143	.12	2.62
M_B-M_I	.1143	.13	.88

TABLE IV

Experiment I

T-TEST OF THE SIGNIFICANCE OF THE DIFFERENCES IN MEAN GAINS IN INTEREST BETWEEN THE NORMAL ORDER (CRITERION) GROUP AND THE EXPERIMENTAL GROUPS

(N = 35)

Orders	Difference	Sigma of difference	C.R.
M_B-M_A	.0286	.07	.41
M_B-M_C	.0000	.08	.00
M_B-M_D	.0857	.09	.95
M_B-M_E	.0000	.08	.00
M_B-M_F	.0571	.09	.63
M_B-M_G	.1143	.09	1.27
M_B-M_H	.0000	.08	.00
M_B-M_I	.0571	.08	.71

Table VIII shows the significance between the normal and random order groups for the questions about the speaker and the speech. This table shows that the random order group liked both the speaker and the speech as well as did the normal order group. However, the normal order group rated the organization of the normal order speech significantly higher than the random order

TABLE V

Experiment I

T-TEST OF THE SIGNIFICANCE OF THE DIFFERENCES IN MEAN GAINS IN ATTITUDE TOWARD HEALTH BETWEEN THE NORMAL ORDER (CRITERION) GROUP AND THE EXPERIMENTAL GROUPS (N = 35)

Orders	Difference	Sigma of difference	C.R.
M_B-M_A	.0888	.11	.81
M_B-M_C	.0857	.11	.79
M_B-M_D	.0285	.11	.26
M_B-M_E	.1714	.12	1.43
M_B-M_F	.2000	.13	1.54
M_B-M_G	.0888	.11	.81
M_B-M_H	.2000	.13	1.54
M_B-M_I	.0857	.11	.79

group rated the random organization. Likewise, the normal order group rated the effectiveness of the normal order speech significantly higher than did those who heard and rated the random organization.

TABLE VI

Experiment I

T-TEST OF THE SIGNIFICANCE OF THE DIFFERENCES IN MEAN GAINS IN ATTITUDE TOWARD MODERATE SOCIAL DRINKING BETWEEN THE NORMAL ORDER (CRITERION) GROUP AND THE EXPERIMENTAL GROUPS (N = 35)

Orders	Difference	Sigma of difference	C.R.
M_B-M_A	.1969	.12	1.64
M_B-M_C	.0571	.13	.44
M_B-M_D	.1428	.13	1.09
M_B-M_E	.1143	.14	.82
M_B-M_F	.2286	.14	1.63
M_B-M_G	.2286	.14	1.63
M_B-M_H	.1143	.12	.95
M_B-M_I	.0571	.13	.44

RESULTS FROM EXPERIMENT II

Examination of Table IX shows that statistically significant changes in mean scores in Experiment II for the normal

TABLE VII

Experiment I

CRITICAL RATIOS OF MEAN GAINS IN SOCIAL ATTITUDE FOR ALL GROUPS

Order	C.R.	Probability of a deviation greater than x	Number of parts transposed
Transitions to Beginning	4.34	.0000	2
Normal Order	3.81	.0001	0
Introduction to Middle	3.23	.0006	1
Random Order	2.99	.0014	7
Introduction and Conclusion Reversed	2.96	.0015	2
Transitions Last	2.51	.0060	2
Conclusion First	2.00	.0228	1
Introduction Last	1.89	.0294	1
Conclusion to Middle	.34	.3669	1

TABLE VIII

Experiment I

T-TEST OF THE SIGNIFICANCE OF THE MEAN DIFFERENCE BETWEEN NORMAL AND RANDOM ORDERS FOR QUESTIONS ABOUT THE SPEAKER AND THE SPEECH (N = 35)

M_B-M_D	Difference	Sigma of Difference	C.R.
Liking speaker	.0286	.124	.23
Liking speech	.0286	.148	.19
Believing speech organized	.6857	.105	6.53
Believing speech effective	.6571	.177	3.71

order (criterion) group occurred for the *social attitude, interest,* and *willingness-to-discuss* questions.

Table X shows the critical ratios of the mean gains for all questions for the various experimental orders. The question on social attitude shows statistically significant chances for both the B (normal order) and the H (introduction to end) group. No other orders show significant gains for this factor.

Table XI shows that the normal order group changed opinions to a greater extent than did any other group. The only other significant shift was made by the group which heard the introduction at the end of the speech. *The random order group changed considerably, but in a direction negative to that urged by the speech.*

Table XII shows that two experimental groups, the D (random) and the E

TABLE IX

Experiment II

GAINS ON THE WOODWARD BALLOT FOR THE NORMAL ORDER (CRITERION) GROUP
(N = 36)

Question	Gain	Mean gain	Sigma Mean	C.R.
1. Importance	4	.1111	.0655	1.69
2. Knowledge	6	.1666	.0735	2.27
3. Social attitude	15	.4167	.1137	3.67
4. Interest	35	.9722	.0273	.3561
5. Willingness to discuss	13	.3611	.0934	3.87

TABLE X

Experiment II

CRITICAL RATIOS OF MEANS GAINS FOR ALL QUESTIONS FOR ALL GROUPS
(N = 36)

Order*	Importance	Knowledge	Social attitude	Interest	Willingness to discuss
B	1.69	2.27	3.67	35.61	3.87
C	1.81	2.50	.63	19.88	2.41
D	2.95	1.15	2.37	19.88	.96
E	1.15	.82	.69	14.95	.00
G	3.72	.00	2.19	24.73	1.81
H	3.04	3.29	3.32	24.73	2.34
I	1.29	1.15	1.11	35.61	3.22

*See order key, page 294.

TABLE XI

Experiment II

CRITICAL RATIOS OF MEAN GAINS IN SOCIAL ATTITUDE FOR ALL GROUPS

Order	C.R.	Probability of a deviation greater than x	Number of Parts Transposed
Normal Order	3.67	.0001	0
Introduction Last	3.32	.0005	1
Conclusion First	2.19	.0143	1
Conclusion to Middle	1.11	.1335	1
Introduction and Conclusion Reversed	.89	.1867	2
Introduction to Middle	.63	.2643	1
Random	2.37*	.0089	6

*This is actually a negative shift. To get the approximate sigma difference between this C.R. and that for the normal order speech, the two critical ratios must be added.

TABLE XII

Experiment II

T-TEST OF THE SIGNIFICANCE OF THE DIFFERENCES IN MEAN GAINS IN SOCIAL ATTITUDE BETWEEN THE NORMAL ORDER (CRITERION) GROUP AND THE EXPERIMENTAL GROUPS (N = 36)

Orders*	Difference	Sigma of difference	C.R.
M_B-M_C	.3334	.174	1.9
M_B-M_D	.7223	.172	4.2
M_B-M_E	.5000	.165	3.0
M_B-M_G	.1667	.161	1.0
M_B-M_H	.0556	.157	.4
M_B-M_I	.2778	.170	1.6

*See order key, page 294.

(introduction and conclusion reversed) had statistically significant differences in mean gain when compared with the normal order group on the *social attitude* question.

CONCLUSIONS FROM THE TWO EXPERIMENTS

Before listing the conclusions of the two experiments it should be pointed out that each was unique and individually complete. The criterion, however, was the same in each instance—the effect of the respective normal order speech. The methodology was the same throughout. The differences were: a. the speech topic, and b. the number of speech parts.

The following conclusions dealing directly with speech organization are listed as primary conclusions; the secondary conclusions are those which became apparent when the correlational relationships among the other factors of the experiments became apparent. This does not necessarily mean that the secondary conclusions are of any less importance than the others.

PRIMARY CONCLUSIONS

1. The transposition of a single main part of a speech to a position in the sequence other than its normal one does not affect the persuasive outcome of the speech. The experimental variations of the two separate experiments, using different topics and different methods of original speech organization, yielded this result without exception. Four experimental transpositions of individual units were made in the first experiment, four similar transpositions were made in the second (Tables VII, XI).

When two parts were transposed, however, the difference between the effectiveness of the resultant speeches and that of the respective normal order speech did prove, in Experiment II, to be significant (C.R. = 3.0).

When all of the parts were transposed, i.e., with complete randomization of the main part order, the second experiment actually yielded a negative audience reaction (Table XI). Thus the evidence indicates that with certain topics and/or methods of total organization in the compositional pattern of the original speech, over-all organization *per se* is an extremely important factor in persuasion.

2. The results of both experiments indicate that statistically significant changes in audience opinion as a result of listening to a short persuasive speech do not occur for:

a. *importance* of a topic (Tables I, IX),
b. belief that *knowledge* about a topic has increased (Tables I, IX).

3. Experiment I showed no significant change for either of the following factors:

a. *Willingness* to discuss the topic (Table I).
b. *personal attitude* toward the practice (Table I).

Experiment II did show a significant increase in *willingness to discuss* the topic (Table IX). The evidence therefore indicates that with certain topics or

methods of organization of the original speech, or both, the audience does become willing to discuss the topic. The *personal attitude* question was omitted from the second experiment.

4. The results of both experiments indicate that statistically significant changes in audience opinion do occur for the normal order speech for:

a. *social attitude* (Tables I, IX).
b. *interest* in the topic (Tables I, IX).

Experiment I also showed significant increases in:

a. belief that a practice is harmful to *health* (Table I).
b. social attitude toward a *moderate* form of the practice (Table I).

These latter two questions were omitted from the second experiment.

5. Modifications of the part sequence of neither of the experimental speeches affected gains in *interest* (Table IV, table omitted). With Experiment I such modifications resulted in no significant differences from the normal order speech for:

a. *social attitude* (Table III).
b. belief that the practice is harmful to *health* (Table V).
c. social attitude toward a *moderate* form of the practice (Table VI).

With Experiment II, however, certain modifications did significantly affect *social attitude* (Table XII).

The questions concerning belief that the practice is harmful to *health* and social attitude toward a *moderate* form of the practice were omitted from Experiment II. Results of the second experiment showed that modifications of the part sequence, in general did not affect opinions of:

a. *importance* of the topic (Table omitted).
b. amount of *knowledge* possessed about the topic. (Table omitted).
c. *willingness to discuss* the topic (Table omitted).

These three latter questions, not having proved significant with the normal order speech in Experiment I were not checked for the experimental variations.

6. The second experiment only showed a highly significant positive correlation in the normal order speech between gain in *social attitude* and *believing the speech to be organized* ($r = .83$).

7. Neither experiment yielded a statistically significant mean difference between the random and normal order groups for the question about *liking* the speaker (C.R. = .23, 2.13). The difference, however, approached significance in the second experiment, which would indicate that the speaker with the better organized speech may be the better liked speaker.

8. Neither experiment yielded a statistically significant mean difference between the random and normal order groups for the question about *liking* the speech (C.R. = .19, 2.09). Again there was evidence in the second experiment that speech organization might affect this factor.

9. The results of both experiments showed that listeners can tell if a speech is well organized (C.R. = 6.53, 3.33).

10. The results of both experiments showed that listeners believe the well organized speech to be the more effective (C.R. = 3.71, 3.10). In the first experiment, however, there was no actual difference in the effectiveness of the two speeches, while in the second experiment the well organized speech really was the more effective.

Secondary Conclusions

1. Both experiments yielded positive correlations between initial status of *social attitude* and each of the following factors:

a. *action* with respect to the respective practice ($r = .42$,* $.19$)
b. *taste* (liking for the practice) ($r = .37$,* $.15$).

Only in the first experiment, however, were these correlations large enough to be significant.

2. Neither experiment yielded significant correlations, with one exception, between gain in *social attitude* and either of the following factors:
a. gain in *interest* ($r = -.11$, $.35$*)
b. *liking* the speaker ($r = .05$, $.25$)

* significant at the 5 per cent level

3. Although the first experiment did not yield a significant correlation between gain in *social attitude* and *liking* the speech, the second experiment did show these factors to be significantly related ($r = .04$, $.78$t).

4. The results of both experiments showed a significant positive correlation between gain in *social attitude* and *believing the speech to be effective* ($r = .37$*, $.66$t).

*Significant at the 5 per cent level
t significant at the 1 per cent level

AN INVESTIGATION IN MEASURING AND IMPROVING LISTENING ABILITY OF COLLEGE FRESHMEN[1]

ARTHUR WILLIAM HEILMAN
College of Education
University of Oklahoma

IN recent years there has been a feeling on the part of some educators that listening is worthy of more attention than it has received and there has been an awakening of interest in listening as a means of learning. Although research in listening has been very limited, there are some problems that have been investigated. These include:

1. The relative effectiveness, in regard to retention, of presenting material aurally or visually.
2. The effect of rate-of-presentation on learning.
3. The study of the relationship between listening ability and other abilities and aptitudes such as intelligence, reading ability and academic success.
4. The effect on learning of presenting material aurally, but using different modes of aural presentation, i.e., lecture, dialogue, story form and dramatization.
5. The effect of aural presentation of material when the speaker is in the presence of the subjects, contrasted with the speaker not being in the presence of the subjects but presenting the material via a loudspeaker.

Objectives of this study:

1. To design a series of training units for improving listening ability.
2. To determine the effectiveness of these units as applied to an experimental group of college freshmen.
3. To determine what effect training in listening would have in listening situations not connected with the training.
4. To build improved instruments, specifically, two comparable forms of a listening test for measuring listening comprehension.
5. To determine the relationship between listening ability, as measured by the listening test used, and intelligence and reading ability.

Procedure:

1. Two comparable forms of a test of listening comprehension, (Forms A and B), were developed for use at the college level.
2. A series of six listening training units were developed to be used at the college freshman level.
3. Two random samples of college freshmen were used. One, the experimental or training group (N-220), was given one form of the listening test as a pre-test, then special training in listening, and finally the other form of the listening test as a criterion test. A control group (N-234), was given a pre-test of listening ability at the same time it was given to the experimental group, but the control group received no special training in listening. This group also took the other form of the listening test at the same time the experi-

[1] Abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate College, State University of Iowa. The author is indebted to Professor James B. Stroud for his assistance and helpfulness throughout the course of the study.

mental group took the post-training test.

4. The means of the two groups on the pre-test were determined and the difference between the means was tested for significance of the difference. The data for both groups on the pre-test of listening are found in Table I.
5. The means of the two groups on the end-test were determined in order to evaluate the effect that training in listening had on the experimental group.
6. To determine if there was any transfer from the training to other listening situations not related to the training, the subjects in both the control and experimental groups were examined over the content of a series of lectures heard in the communication course during the period of time covered by the experiment.

TABLE I

DATA FOR EXPERIMENTAL AND CONTROL GROUPS ON PRE-TEST OF LISTENING COMPREHENSION

	Experimental Group	Control Group	Diff. in Mean
N taking Form A	101	107	
N taking Form B	119	127	
Total N	220	234	
Mean on Pre-Test	52.35	51.59	.76
Range of Scores	36-74	32-73	
Range	38	41	
Standard Deviation	7.63	8.24	
S.E. of the Mean	.51	.54	

Listening Tests Form A and B

Two comparable forms of a test of listening comprehension were built. Each form consisted of three sub-tests. Two sub-tests on each form consisted of items based upon sustained listening material as follows:

Form A. "The Illinois Plan,"[2] a 1,875 word abstract from a pamphlet concerned with education of exceptional children in the state of Illinois, and "The American Scholar,"[3] a 1,604 word passage abstracted from a speech by Howard Mumford Jones, dealing with Emerson's views on American education.

Form B. "Food Preservation,"[4] a 1,997 word passage abstracted from a pamphlet prepared for H. G. Heinz Company, dealing with man's search for improved methods of food preservation and "Education for Freedom,"[5] a 1,679 word passage abstracted from a speech by Robert M. Hutchins, dealing with education in America.

The third sub-test which was common to both forms of the test consisted of two parts, one of which involved critical listening to aurally presented problems dealing with directions and identification of figures. The material was contained in a four page 5½"x8½" booklet provided each subject, although directions and other statements about things to be done with the material were presented aurally.

The second part of this sub-test consisted of a series of stimulus statements each followed by four other statements called test statements. The subject, after hearing the stimulus statement judged

[2] *The Educable Mentally Handicapped,* Circular B, No. 12, Springfield, Ill., 1945.

[3] Jones, H. M., *Problems in Reading and Writing,* Sams and McNeir, Prentice Hall, New York, N. Y., 1949, pp. 357-369.

[4] Swank, E. C., "The Story of Food Preservation," 1943.

[5] Hutchins, R. M., *Problems in Reading and Writing, Sams and McNeir,* Prentice Hall, New York, N. Y., 1936, pp. 338-356.

each of the other four statements as either carrying the same meaning as the stimulus statement or a meaning different from it.

The test of the sustained listening material, sub-tests 1 and 2, forms A and B, was recorded. Records were cut by professional engineers of radio station WKAR, East Lansing, Michigan. The same narrator was used for all the material so that voice quality could not favor one form or the other. Rate of delivery was not held rigidly constant sentence by sentence, but the rate on all records of sustained material fell within the limits of 120 to 140 words per minute. A total of 41 multiple choice questions was prepared to test the listener's comprehension. These were printed and presented to the subjects in test booklets.

The directions, problems, and statements of sub-test 3 were likewise recorded for presentation to the listeners. They worked their responses on a regular IBM answer sheet during short intervals of silence which followed each item. This sub-test also contained 41 items. Thus each form of the test was comprised of 82 items.

Preliminary forms of these tests, consisting of 112 and 117 items, were administered to a sampling of approximately 200 college freshmen for the purpose of securing data for an item analysis, determining reliability and equating forms A and B.

The discrimination index of each item was found by the Flanagan Method. Items not having a discrimination index of .20 and falling within the difficulty limits of 20-80 percent passing, were rejected. The items were divided into three broad intervals on the basis of difficulty and discrimination. The number of items falling into each interval for each category is shown below.

Reliability of the tests was determined, using only the retained items in the computation. First, the reliability of the sub-tests of sustained material was computed by determining the correlation coefficient between forms A and B. The obtained coefficient was .80.

The reliability of the sub-test 3, problems, statements, etc., was computed by determining the correlation between odd and even items and using the Spearman Brown Prophecy Formula to estimate the reliability of the full length sub-test. The reliability obtained was .84.

The Training Program:

Six training lessons were developed and recorded for use with the experimental group. The lessons were presented once a week. Approximately twenty minutes were required for each lesson, ten minutes for the recorded training material, six to eight minutes for answering ten-multiple choice items over the recorded material and three to four minutes for a brief discussion of those questions.

The training material was presented aurally via recordings; test items over this material were in printed form, reproduced by a duplicator. Subjects entered their answers on a regular IBM answer sheet. Each subject used the same answer sheet throughout the training period. Items on the lessons were numbered 1-10, 15-24, 31-40, etc.

The training materials prepared by the writer were built around the following objectives:

	Difficulty			Discrimination		
	20-40	40-60	above 60	20-40	40-60	above 60
Form A	14	38	30	25	38	19
Form B	15	39	28	23	41	18

1. To point out to the student a number of listening habits which authorities believe differentiate between good and poor listeners, such as, good listeners are not passive, they adjust readily to the speaker's plan of organization, they profit from the speaker's summary, they do not react emotionally to the speaker, they do not project their own ideas into what the speaker says, etc.
2. To attempt to build up in the student's mind a respect for listening as a means of acquiring knowledge.
3. To explain "projection" and to show the student how his own ideas color his reactions to speakers and to what speakers say.
4. To provide the student with listening situations in which he can attempt to discover speaker bias, but which at the same time may reveal listener bias.
5. To provide listenting situations in which students can practice recognizing and picking out main ideas.
6. To show the listener how to use facts and details to draw correct inferences.

Results:

The subjects used in this study were college freshmen enrolled in the communication course at Michigan State College. Two random samples were selected, one referred to as the experimental group (N-220), the other as a control group (N-234).

Both groups were given a pre-test of listening ability. Half of each group took Form A and half took Form B. As seen in Table I, the experimental group exceeded the control group slightly but not significantly, the means being 52.35 and 51.59 respectively.

Following the lessons on listening, carried out on the experimental group, both the control and experimental groups were given the alternate form of the listening test referred to as the end-test. On the end-test the experimental group was significantly superior. This difference in the means, favoring the experimental group, was significant beyond the one percent level of confidence. After subtracting the gain registered by the control group from the gain of the experimental group, the difference was still significant beyond the one percent level of confidence. The gain registered by the experimental group over its own pre-test mean was significant beyond the one percent level, while the gain of the control group was not significant. Table II contains the data for both groups on the end-test.

TABLE II
COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS ON END TEST OF OF LISTENING COMPREHENSION

	Experimental Group	Control Group
Mean on End Test	56.37	52.30
Range of Scores	36-72	33-70
Standard Deviation	7.47	8.40
S.E. of the Mean	.50	.55

In order to determine the relationship between gains made subsequent to training and subject's pre-training scores, both the experimental and control groups were divided into four groups based on pre-test percentile rank as follows: 0-25, 26-50, 51-75, 76-100. Mean gains in terms of raw scores, by quarters, are shown on the following page.

It is probable that the gains of the lower quarter of the experimental group were overestimated and those of the upper quarter underestimated due to regression effect. The changes in the control group scores were subtracted from those of the experimental group and only the resultant net gains considered as true gain. Following this pro-

Percentile Intervals	Experimental Group Raw Score Changes	Experimental Group Mean Changes	Control Group Raw Score Changes	Control Group Mean Changes
76-100	55	1.0	(—)80	(—)1.4
51-75	194	3.5	9	.1
26-50	226	4.1	110	2.0
0-25	418	7.6	141	2.5

cedure the mean gains for the experimental group, by quarters from highest to lowest was 2.45, 3.34, 2.11 and 5.05 respectively.

The relationship between listening ability and scores on a general intelligence test was investigated. A correlation coefficient of .56 was obtained between listening scores and general intelligence as measured by total scores on the American Council Psychological Examination.

The relationship between gains following the training and subject's scores on the above intelligence test was studied. The experimental group was divided into 16 groups on the basis of both listing pre-test scores and psychological scores. From an examination of the data found in Table III, it would appear that students whose pre-training listing scores are low in relationship to their standing on the psychological test, profit considerably from the training in listening, as indicated by their gains following training.

Another purpose of the study was to investigate possible transfer effects or generalization of training to situations not directly connected with the training phase of the study. A series of lectures on various phases of communication was presented to both the control and experimental groups, as well as all other students enrolled in the course. Such a lecture was given each week. These lectures were a regularly scheduled part of the communication course.

A test consisting of 36 multiple choice items over some of this material was constructed. It consisted of two subtests, one being administered as a surprise test and the other being a part of the course final examination. A reliability coefficient of .80 was obtained for this test.

The experimental group scored significantly higher on this test than did

TABLE III

RELATIONSHIP BETWEEN GAINS FOLLOWING TRAINING AND SCORES ON PRE-TEST AND PSYCHOLOGICAL TEST

Pre-Test Listening	A.C.E. Psychological Test	N	Mean Change
0- 25 Percentile	0- 25 Percentile	21	7.2
0- 25 Percentile	26- 50 Percentile	18	8.8
0- 25 Percentile	51- 75 Percentile	7	5.0
0- 25 Percentile	76-100 Percentile	1	2.0
26- 50 Percentile	0- 25 Percentile	12	2.3
26- 50 Percentile	26- 50 Percentile	12	5.5
26- 50 Percentile	51- 75 Percentile	11	6.6
26- 50 Percentile	76-100 Percentile	11	5.0
51- 75 Percentile	0- 25 Percentile	9	1.2
51-75 Percentile	26- 50 Percentile	10	4.6
51- 75 Percentile	51- 75 Percentile	13	4.0
51- 75 Percentile	76-100 Percentile	12	4.8
76-100 Percentile	0- 25 Percentile	1	(—)8.0
76-100 Percentile	26- 50 Percentile	3	1.7
76-100 Percentile	51- 75 Percentile	10	0.4
76-100 Percentile	76-100 Percentile	29	1.4

the control group. The difference between the means of the two groups was significant at approximately the one per cent level. Data for both groups are found in Table IV.

TABLE IV

DATA FOR EXPERIMENTAL AND CONTROL GROUPS ON THE TEST OF THE WEEKLY LECTURE

	Experimental Group	Control Group
Mean on Test	23.36	22.32
Standard Deviation	4.10	4.06
S.E. of Mean	.30	.29

The correlation coefficient obtained between listening scores and reading, as measured by the cooperative test of Reading Comprehension, was .66. Different experimenters, using various tests of both listening and reading and working with different age groups, have reported correlations between listening and reading ranging from .39 to .82. Relevant to these discrepancies is a statement by Stroud:[6]

> Problems in learning cannot be solved by averaging the results of a number of separate experiments. One experiment may be said to serve as a check on another experiment only in the event the conditions of the two are equivalent.

Summary and Conclusions:

1. On the basis of the findings of this study, it is concluded that listening ability of college students can be significantly improved through a program of training in listening.

2. There is evidence that improvement in listening habits, resulting from special training, transfer to listening situations not directly connected with the actual training experience.

3. Training in listening, in this study, appeared to be particularly effective with subjects whose listening ability, prior to training, was lower than would be expected on the basis of their scores on a psychological test.

4. The training also appeared to be profitable for low aptitude students, i.e., students in the lowest quarter on the American Council Psychological Test and lowest quarter of the pre-training listening test. A group of 21 such students made a mean gain on the end-test equal to one standard deviation.

5. Considering the amount of time devoted to the training and improvement in listening ability resulting from the training, it is believed that a program of training-in-listening is justified.

6. The correlation found between listening ability and intelligence was .56, between listening and reading .66.

[6] Stroud, J. B., Educational Psychology, Longmans Green Co., N. Y., 1946.

BIBLIOGRAPHY

1. Blewett, Thomas T., "An Experiment in the Measurement of Listening at the College Level," Ed. Doctor's Thesis, University of Missouri, 1949.
2. Brown, J. I., "Why not Teach Listening?" *School and Society*, Vol. 69, (Feb. 12, 1949), pp. 113-116.
3. Corey, S. M., "Learning From Lectures vs. Learning From Reading," *Journal of Educational Psychology*, XXV (Sept. 1934), pp. 459-470.
4. Goldstein, H., "Reading and Listening Comprehension at Various Controlled Rates," Contributions to Education, N. 821. New York: Bureau of Publications, Teacher's College, Columbia University, 1940.
5. Greene, E. B., "The Relative Effectiveness of Lecture and Individual Reading as Methods of College Teaching," Genetic Psychology Monographs, Vol. 4, 1928.
6. Heye, H., "Effectiveness of Selected Auditory Presentations at the Adult Age Level," unpublished Doctor's thesis, University of Iowa, 1941.
7. Hutchins, R. M., "Education for Freedom," *Problems in Reading and Writing*, pp. 338-356, Prentice Hall, New York, N. Y., 1936.
8. "The Illinois Plan for Special Education of Exceptional Children," *The Educable Mentally Handicapped*, Circular Series B, No. 12, Springfield, Ill., 1945.
9. Jones, H. M., "The American Scholar Once More," *Problems in Reading and Writing*, Sams and McNeir, pp. 357-369, Prentice Hall, New York, N. Y., 1949.
10. Larsen, R., "Common and Differential Factors in Reading and Hearing Comprehension," unpublished Doctor's thesis, State University of Iowa, 1938.

11. Loder, E. J., "A Study of Aural Learning With and Without the Speaker Present," *Journal of Experimental Education,* Vol. 6, 1937, pp. 46-60.
12. Nelson, H. E., "The Effect of Rate on the Recall by Radio Listening of Straight News Reporting," unpublished Doctor's thesis, State University of Iowa, 1947.
13. Nichols, R. G., "Factors in Listening Comprehension," *Speech Monographs,* Vol. XV, No. 2, 1948.
14. Russell, R. D., "A Comparison of Two Methods of Learning," *Journal of Educational Research,* Vol. 18, 1928, pp. 235-238.
15. Swank, E. C., "The Story of Food Preservation," Pamphlet prepared for H. J. Heinz Co., 1943.
16. Stroud, J. B., Educational Psychology, Longmans Green Co., New York, N. Y., 1946.
17. Young, W. E., "The Relation of Reading Comprehension and Retention to Hearing Comprehension and Retention," *Journal of Experimental Education,* V (Sept. 1936), pp. 30-39.

AN EXPLORATORY INVESTIGATION OF LIPREADING ABILITY AMONG NORMAL HEARING STUDENTS

JOHN O'NEILL
The Ohio State University

TEACHERS of lipreading face the problems of (1) developing methods of classroom instruction, (2) evaluating the abilities of the prospective lipreading student, and (3) measuring achievement. Little research has accompanied the successes and failures in teaching speech reading. Kitson,[1] Pintner[2] and the Heiders[3] have explored the relationships between lipreading ability and isolated factors, such as the student's reaction time, reaction to form and color, intelligence, and amount of formal education. Their studies were conducted with hard of hearing adults and deaf children. In general their findings were not conclusive. Kitson[4] found that the "synthetic" type of mind coupled with a good memory was necessary for success in lipreading. Pintner[5] found no significant correlation existing between lipreading and intelligence. The Clarke School studies conducted by the Heiders[6] showed a low correlation between educational achievement and lipreading ability.

The present study was based on an assumption that lipreading ability is a skill that is present in some degree among all users of the English language. The study attempted to evaluate the degree of lipreading ability among a group of normal hearing college students and to discover the psychological factors that might be associated with this ability. It was assumed that positive findings would indicate that there is some psychological factor or combination of such factors that can be associated with skill in lipreading. If such a quantitative measurement could be arrived at, it then would be possible to expedite a training program which places major emphasis on these various factors. Such findings might be applied to the instruction of lipreading to the deaf and hard of hearing.

Procedure

Twenty college students who were enrolled in an undergraduate course in speech served as subjects. The mean age of the group was 23 years (range 19-31). All had normal hearing and had never received any lipreading instruction. In fact, this was their first registration in speech although the course was not a performance course. They were given one of the Mason motion picture tests of visual hearing that consisted of ten graded sentences with a possible score of 61 points. Scores from this test were accepted as indications of relative lipreading ability. The scores on the test ranged from 1 to 32 points. The students were administered a series of tests that purportedly explored five personality attributes. The five areas of personality and ability that were explored, and the specific tests that related to each were:

[1] Kitson, H. D., "Psychological Tests for Lip Reading Ability," *Volta Review*, XVII (1915), p. 47.

[2] Pintner, R., "Speech and Speech Reading Tests for the Deaf," *Journal of Applied Psychology*, XIII (1929), pp. 220-225.

[3] Heider F. K. and Grace M. Heider, "Studies in the Psychology of the Deaf," I, *Psychological Monographs* No. 232 (1940), pp. 124-153.

[4] Kitson, *op. cit.*, 47-48.

[5] Pintner, *op. cit.*, 225.

[6] Heider, *op. cit.*, 133-134.

1. Degree of personal adjustment:

a. Rotter Incomplete Sentences:[7] (Semi-structured oral projective test; the subject finishes sentences for which the first word or words are supplied; can be used as a screening device for determining degree of adjustment.)

b. Fisher interpretation of Rorschach:[8] (Measure of rigidity; see 2a below.)

2. Perceptual Abilities:

a. Rorschach:[9] (Oral projective test: ten printed reproductions of bisymmetrical ink blots—five black and white and five colored.)

b. Case-Ruch Test of Spatial Relations:[10] (Paper and pencil test; thirty-two designs that can be completed by assembling the proper parts; measures the ability to perceive rapidly and accurately relationships among objects in space.)

c. Knower-Dusenbury Test of Ability to Judge Emotions:[11] (Paper and pencil test; photographs of male and female during speech and while expressing emotions; eleven items for each speaker.)

d. Sub-tests of Wechsler-Bellevue: (see 4a below.)

3. Reading Ability:

a. Robinson-Hall (Canada) reading test:[12] (Paper and pencil test; measures reading rate and comprehension.)

4. Intelligence:

a. Wechsler-Bellevue Test of Adult Intelligence:[13] (Eleven tests, six verbal and five performance; measures general intelligence as well as verbal and performance intelligence.)

5. Speech Attitude:

a. Knower Speech Attitude Scale:[14] (Paper and pencil test; forty-eight items; measures attitudes toward speech.)

The subjects were ranked in keeping with their scores on the criterion test (Mason film test) and rank order correlations (*rho*) were computed between these rankings and the ones on each of the comparison and sub-tests. The results are summarized in Table I. They indicate that the factors that were sampled in this study have little relationship with lipreading ability. The apparently significant correlation between lipreading ability and performance is maintained in only one of the relevant subtests.

Discussion and Summary

The Heiders[15] in their study of deaf children found very little correlation between lipreading ability, intelligence and educational achievement. They further claimed that good lipreaders reacted to color rather than form. Possibly the differences in educational background of the subjects used, and the type of test materials utilized may account for the difference in findings between the deaf children the Heiders used and the present population of normal-hearing college students.

Several of the correlations, while not significant at the 1% or 5% level of confidence, are of interest. The moderately high negative correlation between

[7] Rotter, J. B., *Manual for Incomplete Sentences* (New York, 1950).

[8] Fisher, Seymour, "Patterns of Personality Rigidity and Some of Their Determinants," *Psychological Monographs* No. 307 (1950), p. 44.

[9] Beck, Samuel J., *Rorschach's Test,* II (New York, 1944).

[10] Case, H. W. and Floyd Ruch, *Manual of Directions, Survey of Space Relations Ability—Form A.* (Los Angeles, 1944).

[11] Dusenbury, Delwin and Franklin Knower, "Experimental Studies of the Symbolism of Action and Voice, Part I; A Study of the Specificity of Meaning in Facial Expression," QJS, XXIV (1938), pp. 424-436.

[12] Robinson, Francis P. and Prudence Hall, *Manual for the Robinson-Hall Reading Tests* (Columbus, 1949).

[13] Wechsler, David, *The Measurement of Adult Intelligence* (Baltimore, 1944).

[14] Knower, Franklin, "A Study of Speech Attitudes and Adjustments," *SM,* V (1938).

[15] Heider, *op. cit.*

TABLE I

CORRELATION (rho) BETWEEN LIPREADING AND COMPARISON TESTS. N, SUBJECTS, 20.

Comparison Test	rho
Rotter Incomplete Sentences	—.02
Robinson Reading Test	
Rate	—.13
Comprehension	—.21
Speech Attitude	—.38
Spatial Relations	.27
Facial Emotions	—.16
Rorschach	
Color	.18
F plus	—.17
Popular	.02
W	.03
Rigidity	.10
Number of Responses	—.05
Wechsler-Bellevue	
General Intelligence	.35
Verbal Intelligence	.18
Vocabulary	.25
Information	—.06
Similarities	.08
Comprehension	.09
Arithmetic	.12
Digit Span	.38
Performance Intelligence*	.55
Block Design	.38
Object Assembly	—.14
Picture Arrangement	.17
Picture Completion	—.28
Digit Symbol*	.47

*Under the assumption that *rho* = *r*, any value that exceeds .43 is significant (5%) and any value that exceeds .55 is highly significant (1%).

Speech Attitudes and lipreading ability is contrary to popular expectations that a proficient lipreader would be "speech oriented." In the area of Performance Intelligence (Wechsler-Bellevue) the significant correlation between scores for Block Design and lipreading performance poses a question as to whether the ability to comprehend relationships (abstracting ability) might not be an area to be further investigated through the use of conceptual tests such as the Hanfmann-Kasanin Test and the Goldstein-Scheerer Color-Form and Objects Sorting Test.

The present study depends for its validity upon the ability of the Mason test to measure lipreading ability. Mason[16] found a high degree of reliability ($r = .92$) for cinematic tests as an instrument for estimating the visual speech comprehension of acoustically handicapped children. The Heiders[17] in an extensive investigation of the skills that are requisite for proficiency in lipreading, employed motion pictures to test lipreading proficiency. They found that such tests were highly reliable. Under the assumption that it does assess this ability either or both conclusions are possible: (1) normal hearing persons do not possess this capacity in any systematic manner, or (2) the ability is unrelated to the abilities that are tested by the personality and intelligence tests used in this study. Conclusion two may be attributable to the fact that skill in lipreading may be of a different nature in different individuals, and such differences cannot be described by a separation of different contributing factors.

A logical extension of the present study would be a further investigation of perceptual skills, i.e., memory span, perceptual field and rapidity of perception, as well as such other personality attributes as rigidity, social consciousness and imagery type. It would be of interest to investigate whether, the good lipreader would be of the visual imagery type as opposed to the auditory, or motor-kinaesthetic type.

This investigation of normal hearing persons does not isolate any skills that relate directly to success in lipreading. A more detailed exploration of lipreading skill is contemplated with major emphasis being placed upon the ability of normal hearing individuals to recognize specific phonemes, words and phrases.

[16] Mason, Marie, "A Cinematographic Technique for Testing More Objectively the Visual Speech Comprehension of Young Deaf and Hard of Hearing Children," Abstracts of Doctoral Dissertations, Ohio State U. (1942).

[17] Heider, *op. cit.*

ENVIRONMENTAL FACTORS DIFFERENTIATING STUTTERING CHILDREN FROM NON-STUTTERING CHILDREN

JOHN PAUL MONCUR
University of California at Los Angeles

Introduction

WHILE authorities have advanced many theories as to the etiology of stuttering, the majority agree that environmental factors are of importance. However, there seems to be little agreement as to what specific, adverse environmental stimuli create the tensions and pressures out of which stuttering may arise.

Fletcher[1] attributes stuttering to the feelings of fear, anxiety, and inferiority that are associated with specific social experiences, and that the memory of these experiences coupled with the pathological emotional response connected with them serve to perpetuate the causal factors. Glasner[2] states that stuttering is not based upon any one combination of environmental factors in that a wide range of stimuli may be associated with this speech defect. Johnson's[3] semantogenic conclusions stress certain specific conditions in a child's environment as being of etiological significance in the genesis of stuttering. Johnson believes that the diagnosis and labeling of a child's non-fluent speech patterns as stuttering, combined with over anxiousness concerning his efforts to speak, give rise to the onset and development of stuttering as a clinical problem. Many psychologists believe stuttering to be a danger signal of an underlying personality disorder—a physical symptom of maladjustment which may be regarded as indicative of inner conflict, frustration, or emotional distress. Thorpe[4] points out that there is a growing tendency, even on the part of former exponents of organic etiology, to ascribe stuttering to disturbances in the psychological life of the child.

Although many investigators have evaluated the causes of maladjustment in children, few, if any, have related the wide range of adverse environmental stimuli specifically to stuttering. A survey of the literature on child psychology yields only meager experimental evidence in this area. Prior to the current trend of research carried on by Johnson and his co-workers, and more recently by Travis and his students, little had been done in this area by speech correctionists. Duncan[5] states that the emphasis has been shifting in the last few years from an analysis of the child to a study of the parent. McCarthy[6] emphasizes that there is a great need for basic psychological research as a means of bringing order out of the chaos which appears to exist in the controversial area of stuttering.

In the meantime, the speech correc-

[1] Fletcher, J. M., The Problem of Stuttering (New York, 1928).

[2] Glasner, P. J., "Personality Characteristics and Emotional Problems in Stutterers Under the Age of Five," *Jour. of Speech and Hearing Disorders*," XIV (1949), 135-138.

[3] Johnson, Wendell, et al, "A Study of the Onset and Development of Stuttering," *Jour. of Speech Dis.*, VII (1942), 251-257.

[4] Thorpe, Louis, *Child Psychology and Development* (New York, 1946), p. 509.

[5] Duncan, M., "Home Adjustment of Stutterers versus Non-stutterers," JSHD, XIV (1949), 255-259.

[6] McCarthy, D., "Language Development in Children," *Manual of Child Psychology*, ed by Leonard Carmichael (New York, 1946), p. 549.

tionist must make a specific application of information amassed by child psychologists and psychiatrists on the maladjusted child, much of which has not been specifically related to stuttering. He must also rely upon information concerning the stutterer's environment compiled by other speech correctionists, much of which is empirical in nature, and which has not been verified by scientific research. Since numerous investigators have urged that more psychological research is needed, it was felt that a specific study of environmental factors which may be related to stuttering would be of value.

The purpose of the present investigation was to determine whether or not there are differentiating factors in selected aspects of the environments of stuttering and non-stuttering children which might be of etiological significance in maladjustment in general and in stuttering specifically.

Procedure

After an extensive survey had been made of the literature concerning environmental factors which might be of etiological significance in creating emotional disturbances in children, the present investigation was organized into the following main divisions: (1) the construction of the questionnaire, (2) the selection of the subjects to be included in the investigation, (3) the administration of the questionnaire, and (4) the tabulation and statistical interpretation of the data obtained from the parent interviews.

In order to give adequate coverage to the important environmental areas that have a bearing on children's behavior, the interview form was divided into main categories which were suggested by writers in the fields of psychology, psychiatry, and speech correction. These areas included selected training areas, selected unfavorable environmental influences, the basic needs of the child, speech development and environment, and the symptomatology of maladjustment. The final form was comprised of 330 questions, each of which was proposed by two or more authorities on child care.

All subjects in the present investigation were made available through the whole-hearted cooperation of the Los Angeles, Burbank, and Beverly Hills School Systems in Southern California. The stutterers' group was comprised of forty-two boys and six girls, ranging in age from sixty-two to ninety-eight months (median: 79.629). After a speech interview, these children were diagnosed as stutterers by the present investigator and at least one other speech correctionist, and in the final analysis, only "severe" stutterers were used in this group. The non-stutterers consisted of forty-two boys and six girls, ranging in age from sixty-five to ninety-five months (median: 80:188). After being carefully matched with a stuttering child according to age, sex, school placement, and (approximately) residential area, each child was interviewed for diagnostic purposes, and was included in the present study only when no stuttering symptoms were observable.[7]

In accordance with school regulations, the previously mentioned school systems arranged personal interviews with the mothers for the present investigator. The length of these interviews varied from two to four hours, and in all cases the interviews were completed in one meeting.

[7] A statistical analysis of the adequacy of the pair-by-pair matching of the children according to age reveals no significant difference between the two groups; the matching according to sex and school placement was complete and exact, and residential area, as was stated, was "approximate."

TABLE I

GROUPS SCORES ON ITEMS CONCERNING SYMPTOMS OF MALADJUSTMENT

Item Number	Item	Number of Correlated Pairs	Number of "Yes" Responses Non-stut.	Stut.	Chi Square
100b	Would you say that your child has definite speech blocks?	48	0	48	48.00**
101	Has your child ever stuttered?	48	3	48	45.00**
56a	Is your child very nervous?	48	6	32	22.53**
56b	Is your child very aggressive?	48	3	24	17.64**
130	Is it necessary to discipline your child very often?	48	6	27	14.23**
111	Is your child a "fussy" eater?	48	12	29	11.56**
79	Does your child have many scraps with other children?	48	11	27	7.53**
79a	Is he the aggressor?	48	8	21	6.26*
126	Does your child have nightmares or night-terrors?	48	11	23	6.00*
59	Does he have any mouth habits? (lip or tongue chewing, etc.)	48	3	14	5.88*
58e	Is he a happy child?	48	48	41	5.14*
90	Does your child quarrel with his brothers and sisters?	40	14	26	4.80*
57	Does your child often refuse to do something you might want him to do?	48	8	17	4.27*
127	Does your child wet the bed?	48	12	21	3.86*
77	Does he daydream quite often?	48	12	21	3.24
132	Has your child had temper tantrums?	48	16	24	2.46
66	Has your child any specific fears?	48	14	21	2.13
56c	Is your child very self-conscious?	48	7	14	2.12
64	Does he weep often?	48	3	9	2.08
61	Has he been subject to vomiting spells?	48	3	8	1.45
62	Did you have any problems with thumb sucking?	48	6	10	.90
56	Is your child very shy?	48	4	8	.90
80	Does he have imaginary playmates?	48	9	5	.63
58	Does he bite his nails?	48	10	14	.50
56d	Is he very melancholy?	48	0	2	.50
78	Does he feel sorry for himself?	48	11	13	.05
63	Does he have nervous spasm or tics?	48	1	2	.00
60	Has your child often been constipated?	48	4	4	.00

**Denotes statistically significant scores at the 1 per cent level of confidence.
*Denotes statistically significant scores at the 5 per cent level of confidence.

Since the investigation involved the use of pair-matched groups, a chi square test for correlated samples was applied to the data.[8]

FINDINGS AND DISCUSSION

Symptomatology of Maladjustment

Information was requested of the mothers concerning various aspects of their child's behavior which might possibly indicate maladjustment. Of the twenty-eight items asked on this section, seven differentiated the two groups at the 1 per cent level of confidence and seven at the 5 per cent level. The results on this section are tabulated in Table 1.

According to the mothers' reports, every stuttering child in the present study exhibited several symptoms of maladjustment, which findings seem to substantiate those of Glasner,[9] who reports that all of the young stutterers in his study exhibited symptoms of emotional disturbance other than the stuttering itself. The parallel between the findings of the Glasner study and the present investigation seems to suggest that stutterers do display more symptoms of maladjustment than do nonstutterers and that it is not uncommon for stutterers to display several symp-

[8] For information regarding the use of chi square for correlated samples, cf. McNemar, Q., *Psychological Statistics* (New York, 1949).

[9] Glasner, *op. cit.*, p. 137.

toms, particularly those of general nervousness, problem behavior, enuresis, and nightmares.

A possible explanation of the findings in this section would be that stuttering is merely a symptom of basic maladjustment, of which there are other demonstrable symptoms, several of which were uncovered in this study. Inasmuch as the stutterers were found to exhibit several other symptoms of maladjustment, one might seriously consider that the stuttering as well as the other symptoms resulted from the same tensions, pressures, and frustrations. One might also consider the possibility that some of these other symptoms may have resulted from the stuttering, since the child who cannot communicate adequately becomes frustrated and often resorts to other unsuccessful methods of adjustment.

Selected Training Areas

Discipline. Of particular interest to the present investigation was the finding that 55 per cent of all items asked pertaining to discipline recorded significant differences between the two groups of mothers. The fact that almost all child care authorities feel that unfavorable methods of discipline may be a source of maladjustment in the child tends to emphasize the importance of the findings of this area of parental control.

It was found that significantly more mothers of stutterers than mothers of non-stutterers reported employment of harsh disciplinary measures to control their child's behavior, namely, the use of threats (chi square: 16.67), shame or humiliation (chi square: 10.70), and corporal punishment (chi square: 6.75); all of the aforementioned chi squares exceed that required for the 1 per cent level of confidence. Harsh forms of discipline in themselves may not precipitate stuttering, but, rather, serve as factors in creating emotional instability in the child. Certainly the methods mentioned above are not desirable. Child care authorities generally agree that corporal punishment is a poor method of control, and if practiced to any extent, may have serious repercussions in the child. Shame and humiliation are both considered unfavorable factors in that this method of punishment is a direct blow to the child's ego. On the remaining items dealing with specific disciplinary measures, no significant differences are noted between the two groups on such methods as scolding, removing material things or taking away privileges, isolation, reasoning, and ignoring the child. It is interesting to note that these methods of discipline are considered less severe forms than are those discussed which were found to differentiate the two groups.

That the parents of the stuttering children (as a group) showed very little consistency in discipline is indicated by many of the numerous direct and indirect questions which were felt to reveal a consistent policy. Significantly more mothers of stutterers than of non-stutterers reported that they had failed to keep their disciplinary measures (in general) consistent (chi square: 11.08; $P<.01$), they had been more severe on one occasion than another in handling the same disciplinary problem (chi square: 7.20; $P<.01$), that their child bothered them more on some days than on others (chi square: 5.26 $P<.05$), and that their personal feelings had noticeably altered their control measures (chi square: 5.26; $P<.05$). The parents who do not employ any consistent method of handling their child's behavior tend to create confusion in the child's understanding of the rules that govern his con-

duct, which confusion may create emotional upheavals and frustrations.

Conflict in authority is closely related to inconsistency, for when parents cannot agree on disciplinary measures, no consistent method of control can be established. Significantly more mothers of the stuttering children than of the non-stuttering children stated that they did not support their husband's disciplinary measures (chi square: 6.76; $P<.01$), and he, in turn, did not support theirs (chi square: 8.17; $P<.01$). Parental dissension concerning authority not only confuses a child, but often results in no authority whatsoever.

Significantly more mothers of stutterers than of non-stutterers reported that they reminded their child constantly to sit up straight (chi square: 6.55; $P<.05$), to put away his toys (chi square: 5.54; $P<.05$), and to keep his feet off the furniture (chi square: 4.55; $P<.05$). These replies were felt to indicate more over-supervision and nagging on the part of the mothers of the stutterers. More problem behavior was encountered by the mothers of the stuttering children than by the mothers of non-stuttering children in that more mothers of the former group reported that it was necessary to discipline their child "very often" (chi square: 14.23; $P<.01$). This finding seemed to be corroborated somewhat by the responses to the question asking whether the mothers felt that their child was as well-behaved and as advanced (socially) for his years as he should be, wherein significantly more mothers of stutterers than of non-stutterers evaluated their child's conduct negatively (chi square: 10.32; $P<.01$). Since the response to this question represents a parental judgment rather than a fact, one must keep in mind the possibilty that the differences in response may actually be attributed to a difference in the parents rather than in the child's behavior. Although the parents are expressing an opinion rather than stating a fact, the opinion or attitude may serve as an operating factor upon the child whether the belief is justified or not.

In attempting to solve the enigma of whether emotional disturbances precede stuttering or whether the reverse is true, one may conclude that since stuttering appears after the first few years of life, it does not seem likely that the parents became injudicious and inconsistent because of the stuttering in their child, but, rather, the parents were injudicious and inconsistent in administering their disciplinary methods prior to the onset of this speech disorder; therefore, if the findings are valid, one cannot overlook the possibility that undesirable measures of parental control are a contributing factor to general emotional disturbances out of which stuttering may develop.

Eating habits and training. The mothers of both groups were asked fifteen questions concerning eating habits and training of the children included in the present investigation, five of which reveal significant differences in response between the two groups.

When asked whether or not their child was a "fussy" eater, twenty-nine mothers of stutterers and twelve mothers of non-stutterers said that he was (chi square: 11.56; $P<.01$). Most authorities agree that, as a rule, a child will not become a feeding problem if his parents do not make an issue of eating. Significantly more mothers of stuttering children than of non-stuttering children rewarded their child for eating rejected food (chi square: 9.00; $P<.01$), gave their child food which he desired when he had rejected his meal (chi square: 8.91; $P<.01$), allowed him

to eat in-between meals when he would not eat his dinner (chi square: 8.05; $P<.01$), and did not allow him to feed himself at an early age (chi square: 5.88; $P<.05$). These pampering practices place an undue emphasis on eating, which might possibly disturb the child's ability and willingness to accept food, thus creating a "fussy" eater and an eating problem. It has long been known that the amount of stuttering of an individual varies considerably, depending upon certain factors, of which health is one. Eating problems have been known to cause increased irritability and poor health conditions, which may ultimately be reflected in a child's speech fluency. Since eating habits and training are established before the onset of speech and stuttering, it may be that unfavorable procedures in this regard serve as antecedent causal factors of the general emotional disturbance out of which stuttering may arise.

No significant differences are apparent in this area regarding breast feeding, use of special menus, eating rejected food at the next meal, and regularity in eating hours.

Sleep habits and training. Problem behavior and disciplinary action appeared as major points of difference between the two groups on the phase of the investigation devoted to sleep habits and training. Significantly more mothers of stutterers than of non-stutterers reported that their child refused to go to bed often (chi square: 13.76; $P<.01$). When issues were made concerning the child's retiring at the designated hour, significantly more parents of stutterers than those of non-stutterers employed coercive measures, such as threatening (chi square: 9.80; $P<.01$), scolding (chi square: 8.00; $P<.01$), coaxing (chi square: 5.14; $P<.05$), and spanking (chi square: 4.50; $P<.05$), to make the child obey. These practices seem to indicate a specific application of the methods reported as being employed to a greater extent by the parents of the stutterers in the section on discipline, and in a sense serve as a separate check on the mothers' verbal reports.

Although the parents of the stutterers were more extreme in their methods of control regarding sleep habits, they were also more indulgent inasmuch as significantly more mothers of stutterers than of non-stutterers stated that they sometimes capitulated to their child's desire to remain awake (chi square: 7.58; $P<.01$).

The practice of insisting upon their child going to bed "immediately" without allowing him to finish what he was doing was employed by more mothers of stutterers than mothers of non-stutterers (chi square: 8.10; $P<.01$). The extensive use of this bed-time policy might possibly explain the greater amount of conflict encountered by the mothers of the stuttering children in getting their child to go to bed. There is a certain element of frustration affixed to being torn away from an interesting pursuit without an opportunity to complete the activity, and most authorities are in agreement as to the undesirable aspects of this procedure.

If these reports are valid, then certainly the children in the stuttering group suffered greater emotional disturbances prior to retiring than did the non-stuttering group, which may be considered another factor contributing to the general emotional instability of these children. On the other hand, one must recognize the possibility that the emotional disturbances may be caused by the stuttering in that stuttering may cause the child to be highly nervous and unstable and subject to displaying

emotional outbursts with little provocation. It does not, however, account for the methods employed by the parents in handling the situation once it had occurred.

No statistically significant differences between the two groups are apparent with regard to the following phases of sleep habits and training: punishment for not wetting the bed, parental permission to take toys to bed, number of children told bed-time stories, number of children sent to bed as a form of discipline, number of children possessing private beds and bedrooms, methods of handling the child when he cried at night, participation in games of activity before going to bed, and the number of children reported by their mothers as not being able to sleep without special equipment.

Toilet and sex training. In comparison with certain other aspects of the present investigation, the questions related to toilet and sex training do not reveal any great disparity in response between the two groups. Many parents had difficulty in answering questions pertaining to toilet training in that several years had passed since they had been confronted with this problem. No differences were apparent between the two groups with regard to the various methods employed in toilet training, with the exception of hurrying the child, in which case significantly more mothers of stutterers than of non-stutterers reported that they had attempted to perfect this training in a relatively short period of time (chi square: 5.79; $P<.05$). Of the twelve questions asked pertaining to sex training, only one differentiated the two groups significantly, namely, more mothers of non-stutterers than of stutterers reported that their child had asked about matters concerning sex (chi square: 7.35; $P<.01$). No differences were noted on items dealing with explaining the facts of life or methods employed by the parents in handling shocking behavior. In view of the findings on toilet and sex training, the present investigator cannot conclude that these areas were a source of undue pressure in the children studied.

Selected Unfavorable Environmental Influences

According to the majority of psychologists, there are certain negative conditions, which, if present in the home environment, may be the source of pressures, tensions, and conflicts in the child, ultimately resulting in maladjustment.

Parental friction. Questions asked concerning parental friction reveal that significantly more mothers in the stuttering group than in the non-stuttering group reported unsatisfactory relations with relatives (chi square: 7.58; $P<.01$.), conflict with relatives (chi square: 6.55; $P<.05$), arguments that destroyed the peace of the family (chi square: 8.33; $P<.01$), and parental friction concerning discipline (chi square: 6.76; $P<.01$). No differences are apparent between the two groups regarding parents sharing friends in common and conflict over friends. That the child reared in a happy, companionable home which is relatively free from discord has a much better opportunity to become a well-adjusted personality than the child who is reared amid inharmonious surroundings is a generally accepted belief of almost all child psychologists; thus, the differences found to exist between the two groups under observation in regard to parental friction are very important findings. The findings in this area seem to substantiate Baruch's[10] conclusions

[10] Baruch, Dorothy, "A Study of Reported Tension in Interparental Relationships as Co-Existent with Behavior Adjustment in Young Children," *Jour. of Exper. Educ.*, VI (1937), 187-204.

in her study that parental friction over discipline and relatives was a contributing factor to maladjustment in children. Since Baruch and other psychologists have found that parental friction may cause emotional disturbances and maladjustment in children, and since emotional disturbances are known to cause increased stuttering, one might conclude from the findings that parental friction may operate as a cause (or aggravation factor) in stuttering.

Sibling rivalry. A number of questions were asked which were felt to reflect the presence of sibling jealousies. According to the mothers' testimonies, more children in the stuttering group than in the non-stuttering group were dominated by a brother or sister (chi square: 10.71; $P<.01$), the children quarreled (more than normal) with siblings (chi square: 4.80; $P<.05$), and more unfavorable comparisons were made by the parents (chi square: 5.76; $P<.05$). Several items, such as position of the child in the family, parental favoritism, lack of attention because of ordinal position in the family, wearing clothes that belonged to other children, etc., reveal no statistical differences between the two groups. The greater indication of sibling jealousies in the stuttering group suggests a possible source of tension and conflict, which may be regarded as an aggravation factor in stuttering in that increased stuttering is often associated with increased emotional tensions in the child. In order to determine the extent and effect of sibling rivalry present in both groups, the writer feels that it would be necessary to question both the child and his siblings, since the information amassed from the parents presents a rather one-sided appraisal of sibling conditions.

Excessively high standards. On various direct and indirect questions pertaining to excessively high standards being held for the child it was noted that significantly more mothers of stutterers than mothers of non-stutterers felt that their child had to struggle to keep up to the high standards set by his family (chi square: 21.55; $P<.01$), that their child had been punished or shamed for a poor speech performance (chi square: 12.07; $P<.01$), that their child was not as advanced (socially) nor as well-behaved for his years as he should be (chi square: 10.32; $P<.01$), and that their child's toilet training had been attempted in a relatively short period of time (chi square: 5.79; $P<.05$). No differences are revealed on questions concerning teaching the child to read and write before going to school, stressing the importance of good grades, and expecting more of the child because of his size. Many psychologists believe that to push a child beyond his capacity puts him under a constant strain and creates nervousness. The findings of this investigation would seem to corroborate this concept as two of the greatest differences established in the entire study were on too-high standards being held (chi square: 21.55) and reported nervousness (chi square: 22.53). It was also noted that more mothers of stutterers than those of non-stutterers indicated dissatisfaction with their child's immature conduct, which, although an attitude, might possibly reflect too-intense urging of the child by the parents to attainments beyond his reach by legitimate means, thereby contributing to the problem behavior with which they showed displeasure.

Illness in the family. An analysis of the information obtained on illness or health problems in the family reveals that significantly more mothers of stutterers reported serious illness, accidents, or health conditions which had affected

the household than did the mothers of non-stutterers. As to the effect of illness in the home upon the children under investigation, one can only speculate, for when questioned concerning this, most mothers said that they really did not know. Authorities, however, indicate that extended illness of a serious nature within the home may cause feelings of fear, insecurity, and anxiety in the child, which, if operative, may serve as grounds for greater instability and possibly increased non-fluency in the child.

Presence of extra adults in the family. On questions regarding the presence of extra adults in the family, no significant differences were found between the two groups. In several cases, mothers reported that chaotic home conditions had been created by this factor, but since both groups reported such conditions, no difference could be established.

The broken home. An analysis of the data obtained concerning whether or not the home had been broken by death, separation, or divorce reveals no significant difference between the two groups, nor was the difference significant with respect to the item asking whether or not a parent had been separated from the child because of being in the army, or for job or illness reasons.

Speech Development and Environment

A large number of questions were asked regarding the speech development and environment of the children included in the investigation. Detailed information concerning these questions is presented in Table II.

The majority of speech correctionists agree that to correct a child when he is having difficulty with his speech may lead to more serious speech disturbances. Frequent correction also suggests parental over-supervision, fault finding, and nagging, which, warns Dreikurs,[11] discourages most children, impedes their conduct and achievement, and promotes disobedience and failure. On questions pertaining to the correction of their child's speech, significantly more mothers of stutterers than those of non-stutterers reported that they had supplied the word, told him to "stop and start over," called his attention to his speech, scolded him, and told him to "think before he talked." If the hypothesis of Johnson and his followers (that what the layman diagnoses and reacts to as stuttering is the normal non-fluent speech of young children) is valid, then the above correction factors are of primary importance in the onset and development of this speech disorder. Why the parents of stuttering children are over-anxious and demanding concerning their child's speech while parents of non-stuttering children are not is not adequately explained in the Johnsonian theory. One explanation may be that the parents of stutterers have certain dominating characteristics which the parents of non-stutterers do not possess, at least insofar as speech is concerned. On the other hand, it may be that the speech patterns of the stuttering children are actually more hesitant and repetitious than are those of the non-stutterers, which extreme non-fluency is corrected in the manner described earlier. Why some children are more non-fluent than others appears to be a relatively unexplored area of research. It may be that the greater number of adverse environmental factors found to exist in the stuttering group than in the non-stuttering group served as causal factors in creating greater non-fluency in the stuttering group, which

[11] Dreikurs, R., *The Challenge of Parenthood* (New York, 1948).

TABLE II

GROUP SCORES ON ITEMS CONCERNING SPEECH DEVELOPMENT AND ENVIRONMENT

Item Number	Item	Number of Correlated Pairs	Number of "Yes" Responses Non-stut.	Stut.	Chi Square
102b	When your child had difficulty with his speech, did you call his attention to it in any way?	48	5	28	21.16**
102g	Did it seem to disturb your child when you corrected him?	48	1	24	21.16**
102a	When your child had difficulty with his speech, did you tell him to "stop and start over?"	48	15	40	18.94**
99	Has anyone made fun of your child's speech?	48	6	32	18.78**
105	Has anyone ever called your child a stutterer?	48	0	16	14.06**
103	Is there anyone in your family or your husband's family who stutterers?	48	7	23	12.80**
102	When your child had difficulty with his speech, did you supply the word?	48	13	31	12.46**
96	Has anyone punished or shamed your child for a poor speech performance?	48	0	14	12.07**
105a	Did it cause your child to be disturbed when he was called a stutterer?	48	0	13	11.08**
102c	When your child had difficulty with his speech, did you scold him?	48	1	13	8.64**
102e	When your child had difficulty with his speech, did you tell him to "think before he talked?"	48	18	31	6.76**
98a	Do you feel this has created unequal verbal competition? (see item 98)	48	1	8	4.00*
94b	Do you encourage your child to go to the door and answer the telephone?	48	43	35	3.06
93	Does your child have difficulty in making himself understood?	48	10	17	2.42
102d	When your child had difficulty with his speech, did you ignore it?	48	44	39	1.45
92a	Did your child have difficulty with special sounds?	48	16	11	1.19
54c	Did he ever seem to use both sides equally well? (see item 54)	48	14	19	1.19
92	Did your child have any particular speech difficulties at an early age?	48	6	11	1.07
95	Does your child recite nursery rhymes?	48	42	38	.90
54a	Was this his own choice? (see item 54)	48	45	41	.90
97	Do you reward your child for speaking well?	48	4	1	.80
95a	Does your child say nursery rhymes in front of people?	48	22	18	.73
92b	Did you try to help him with special speech sounds?	48	15	11	.64
94	Does your child convey messages over the telephone?	48	43	39	.44
98	Does your child have a brother or sister who talks better than he does?	48	5	8	.31
93a	Did you try to "ease the way for him"? (see item 93)	48	10	13	.26
103a	Did your child ever imitate a stutterer?	48	5	3	.17
94a	Does your child go to the door?	48	43	45	.17
95b	Does your child perform of his own free will?	48	41	39	.08
104	Has your child stuttered to attract attention?	48	2	3	.00
54	Is your child right-handed?	48	41	42	.00
54b	Was he ever switched from one side to the other?	48	5	5	.00

**Denotes statistically significant scores at the 1 per cent level of confidence.

*Denotes statistically significant scores at the 5 per cent level of confidence.

was subsequently diagnosed and reacted to as stuttering.

Significantly more mothers of stutterers than of non-stutterers shamed or humiliated their child for a poor speech performance. In resorting to this method of correction, a parent is not only calling the child's attention to the fact that his speech is abnormal, but is helping to create feelings of inferiority, guilt, and inadequacy concerning it as well. It is interesting to note that shame and humiliation were used by more mothers of stutterers than of non-stutterers as a method of control in general (cf. *Discipline*), which would seem to corroborate the above finding.

The findings on items pertaining to laterality appear to contradict earlier research studies reporting these factors as being of importance in the onset and development of stuttering, as information obtained regarding handedness, choice of hand, switching of handedness and ambidexterity of the children studied do not reveal any significant differences between the two groups.

The Basic Needs of the Child

An analysis of the basic needs of the child was undertaken in order to determine whether those conditions which a child is believed to need physically, mentally, and socio-economically in order to live effectively and happily were fulfilled to a greater extent in one group than in the other, and whether any relationship between stuttering and the lack of fulfillment of these needs could be established.[12]

The basic need for feelings of security. Fenton[13] stresses that no other factor is more significant psychologically to the child than is the feeling of security. The findings that significantly more mothers of non-stutterers than of stutterers reported that their family was a "closely-knit" unit (chi square: 12.80; $P<.01$), and that the family shared experiences together (chi square: 7.56), laughed a lot together (chi square: 13.47), went on vacations together (chi square: 12.46), went to shows together (chi square: 6.72), and listened to programs together (chi square: 6.72), suggest that the stuttering children (as a group) were not exposed to as many happy family experiences as were the non-stuttering children. Carrying on activities together and developing mutual interests do much toward establishing good parent-child relationships, which in turn contribute to the child's feelings of security—a factor which may account, in part, for the greater speech stability of the non-stutterers. Symptoms of anxiety, such as worry, nervousness, enuresis, and nightmares were reported for more stutterers than non-stutterers, which symptoms, according to Fenton, are indicative of feelings of insecurity in the child. Greater economic security in the non-stuttering group was suggested by the findings that more mothers in this group than in the stuttering group reported that they owned their own home and that they had not expressed financial worries in front of the child. No differences between the two groups are apparent with regard to time spent with the child, number of places lived, insecurity in husband's work, and ownership of various kinds of material possessions, such as radios, television sets, automobiles, furniture, etc.

The basic need for a healthy body and good physique and appearance. The findings of this basic need may be summarized as follows: more mothers of

[12] The Basic needs did not appear as separate sections in themselves; rather, they cut longitudinally through the many divisions of the investigation.

[13] Fenton, Norman, *Mental Hygiene in School Practice* (Stanford, 1943), p. 185.

stutterers reported that their child did not have good coordination (chi square: 8.10; $P<.01$), had a normal birth (chi square: 4.55; $P<.05$), and had undergone a severe emotional experience (chi square: 3.85; $P<.05$). The large bulk of questions on this area, however, showed no differences between the two groups, particularly those dealing with having physical defects or weakness, playing well for age, being large or small for age, being unduly concerned when hurt or ill, and the number and severity of childhood illnesses.

The basic need for social adjustment and recognition, and the need to accept the conditions and realities of life. The need to accept the conditions and realities of life is closely associated with the need for social adjustment and recognition, for if one cannot adjust himself to the actual conditions of his life, recognizing the reality and the inevitability of the conditions to which he must adjust, he cannot adequately adjust himself to others. On these phases of the investigation, it was found that significantly more mothers of non-stutterers than of stutterers reported that their child enjoyed school (chi square: 9.39; $P<.01$), had many friends at school (chi square: 6.75; $P<.01$), got along well with his playmates (chi square: 10.08; $P<.01$), and often played with his siblings (chi square: 7.56; $P<.01$). On the other hand, more mothers of stutterers than of non-stutterers stated that their child often quarreled with siblings (chi square: 4.00; $P<.05$), got in many "scraps" with children (chi square: 7.53; $P<.01$), was usually the aggressor in these fights (chi square: 6.26; $P<.05$), and displayed evidences of negativistic behavior (chi square: 4:27; $P<.05$). The findings in these two areas of basic needs not only tend to substantiate one another, but also appear to substantiate the belief held by many speech pathologists that stutterers on the whole are not adjusted socially and often do not accept the conditions and realities of life.

The basic need for feelings of competence. In analyzing the data pertaining to this basic need, these major points of difference appear to exist between the two groups. Significantly more non-stutterers than stutterers were reported by their mothers as owning pets (chi square: 6.00; $P<.05$), taking care of them personally (chi square: 7.20; $P<.01$), having special interests or hobbies (chi square: 7.20; $P<.01$), sustaining interest in new activities or hobbies (chi square: 6.72; $P<.01$), and doing satisfactory work in school (chi square: 5.82; $P<.05$)—all of which indicate a greater degree of fulfillment of this basic need in the former group than in the latter. No differences are apparent, however, with regard to owning and riding bicycles in the street, being fearful or eager to try new activities, enjoying study, reading well for age, or being encouraged in new pursuits.

The basic need to experience curiosity and pleasure and to acquire active and varied interests. A large number of questions were devoted to exploring the various activities of the children under investigation, namely, his equipment and play space, the organizations to which he belonged, his hobbies and special activities, trips that he had made, etc. An analysis of the data amassed on this phase of the investigation reveals that the two groups were approximately equal with regard to their activities and play equipment.

The basic need to be considered a developing personality. This basic need is closely associated with a number of other basic needs. Usually the child who has a feeling of competence has been treated

as a developing personality; also, the child as a developing personality needs to have feelings of security, which are based upon the belief that his parents, friends, and teachers have faith in his ability to adjust to life's conditions. With regard to this area it was found that more mothers of non-stutterers than of stutterers assigned their child to definite responsibilities around the house (chi square: 7.59; $P<.01$), and made this a definite routine (chi square: 16.67; $P<.01$). The findings previously reported that more non-stutterers had pets and were allowed to take care of them also suggest that more responsibility on the whole was given to the children in this group. A possible explanation of these findings may be that the parents of the stutterers had a tendency to over-protect or "baby" their child, feeling that he was as yet incapable of assuming responsibility. On the other hand, no differences between the two groups are noted with regard to running errands of various kinds, going to shows alone, having allowances, and owning various kinds of equipment.

Conclusions

On the basis of the information obtained from the mothers, one may conclude that the stutterers, as a group, were subjected to more adverse environmental influences than were the non-stutterers, and that they were noticeably disturbed by these factors in that they displayed more symptoms of general maladjustment other than the stuttering itself than did the children in in the non-stuttering group. Whether the wide range of environmental pressures brought to bear upon the stutterer preceded the child's stuttering and was a contributing factor to this speech disorder or whether the child's frustration because of his speech problem precipitated deviate forms of behavior which evoked these pressures can be determined to an extent by the findings of the present investigation. Ample evidence has been presented that adverse environmental influences were operative prior to the onset of stuttering in the children of the stuttering group. For example, the training procedures investigated, such as discipline, eating, sleeping, and toilet training are concerned with the earliest years of the child's life, and the significant differences which were found to differentiate the two groups in these areas of the investigation can hardly be attributed to any speech problem exhibited by the child. Likewise, it is extremely doubtful that the children's stuttering accounted for the greater evidences of parental friction, unhappy home conditions, excessively high standards, and sibling rivalry which were indicated to a greater extent in the stuttering group than the non-stuttering group.

Since many of these adverse environmental stimuli appear to have existed prior to the onset of stuttering in the children of the stuttering group, and since almost all psychologists believe that these influences give rise to emotional instability and subsequent maladjustment in children, it would seem to follow that the stutterers suffered emotional disturbances before the precipitation of their speech disorder. If specific pressures brought to bear upon a child with regard to his speech can cause serious impairment in the speech flow (Johnson's diagnosogenic and semantogenic conclusions), it does not seem unlikely that emotional disturbances in general could also influence the stability of the speech process. Such speculation appears to answer in part the question raised earlier as to why one group

of parents diagnosed and labeled their child's speech as stuttering while the other did not.

As to whether or not there exists a syndrome of environmental factors which precipitates stuttering with regularity, one may only speculate. If there exists one causal factor which may account for the manifold differences discovered between the stuttering and non-stuttering groups and to which all other environmental factors are subordinate, it may be that the parents of the stuttering children are (as a group)—dominant parents. It may be recalled that significantly more mothers of stutterers than of non-stutterers used harsh forms of discipline, held their child to excessively high standards, evaluated their child's behavior negatively, and over-supervised their child—all of which according to Symonds and Fitz-Simmons, are characteristic of the dominant parent.

In summarizing theoretical considerations in the present investigation, the following hypothesis is offered: what parents and others often diagnose as stuttering in children are the exaggerated non-fluent speech patterns of a frustrated and confused child overwhelmed by a wide variety of adverse environmental stimuli, and that these stimuli which also precipitate general emotional disturbances are the antecedent causal factors of the exaggerated interruptions in the child's speech flow, which are subsequently diagnosed and reacted to as stuttering. This evaluation and labeling of the child's speech pattern together with a continuation of the unfavorable environmental influences tend to fixate and perpetuate the syndrome of stuttering as it is found in the older child and adult.